JOURNAL OF CONSULTING PSYCHOLOGY

Edited By
Edward S. Bordin
University of Michigan

Advisory Editors

John W. Atkinson
University of Michigan

John M. Butler
University of Chicago

Seymour Fisher
Baylor University

Harrison G. Gough
University of California

Robert B. Malmo
McGill University

Boyd R. McCandless
University of Iowa

Paul E. Meehl
University of Minnesota

Daniel R. Miller
University of Michigan

Ruth L. Munroe
New York, N. Y.

Anne Roe
New York University

Saul Rosenzweig
Washington University

Julian B. Rotter
Ohio State University

Nevitt Sanford
University of California

Seymour B. Sarason
Yale University

Robert R. Sears
Stanford University

Edwin S. Shneidman
Veterans Administration, Los Angeles, California

Read D. Tuddenham
University of California

LIBRARY

VOLUME 23

1959

Arthur C. Hoffman, *Managing Editor;* Helen Orr, *Promotion Manager;*
Barbara Cummings, *Editorial Assistant*

Published Bimonthly By

The American Psychological Association, Inc.
1333 Sixteenth St., N.W.
Washington 6, D. C.

Printed at
Lancaster, Pennsylvania

$8.50

Copyright © 1959 by the American Psychological Association, Inc.

Bureau of Edl. & Psyl. Research
(S. C. E. R. T.)
Date 6·3·81
Acc. J693

Contents of Volume 23

Adams, Abby B. Choice of Infant Feeding Technique as a Function of Maternal Personality 143

Affleck, D. C., and Mednick, S. A. The Use of the Rorschach Test in the Prediction of the Abrupt Terminator in Individual Psychotherapy - - - - - - - - - - - - - 125

Akel, M. See *Estes, Betsy W.*

Allison, J., and Hunt, D. E. Social Desirability and the Expression of Aggression under Varying Conditions of Frustration - 528

Anderson, A. C. See *Peterson, D. R.*

Angelino, H. See *Nebergall, Nelda S.*

Astin, A. W. A Factor Study of the MMPI Psychopathic Deviate Scale - - - - - - 550

Baron, M. R. See *Hoyt, T. E.*

Bauer, R. W. See *Johnson, D. E.*

Becker, W. C., Peterson, D. R., Hellmer, L. A., Shoemaker, D. J., and Quay, H. C. Factors in Parental Behavior and Personality as Related to Problem Behavior in Children - - 107

Bendig, A. W. Score Reliability of Dichotomous and Trichotomous Item Responses on the Maudsley Personality Inventory - 181

Bendig, A. W. "Social Desirability" and "Anxiety" Variables in the IPAT Anxiety Scale 377

Beier, E. G., and Stumpf, J. Cues Influencing Judgment of Personality Characteristics - 219

Bernstein, L. See *Pile, E.*

Bills, R. E. Two Questions: A Reply to Cowen and Tongas - - - - - - - - - - - 366

Bindman, A. J. Mental Health Consultation: Theory and Practice - - - - - - - - - 473

Bloom, B. L. Ecologic Factors in the WAIS Picture Completion test - - - - - - - - 375

Bordin, E. S. Editorial - 1

Bostian, D. W. See *Lasky, J. J.*

Bridges, C. Nomographs for Computing the "Validity" of WISC or Wechsler-Bellevue Short Forms - 453

Brown, D. R. See *Johnson, D. E.*

Buss, A. H. The Effect of Item Style on Social Desirability and Frequency of Endorsement 510

Caditz, S. B. Effect of a Training School Experience on the Personality of Delinquent Boys 501

Calden, G., Lundy, R. M., and Schlafer, R. J. Sex Differences in Body Concepts - - - - 378

Cameron, G. R. See *Peterson, D. R.*

Campbell, D. T. See *Chapman, L. J.*

Capell, M. D., and Wohl, J. An Approach to the Factor Structure of Clinical Judgments 51

Cattell, R. B. See *Peterson, D. R.*

Chapman, L. J., and Campbell, D. T. Absence of Acquiescence Response Set in the Taylor Manifest Anxiety Scale - 465

Chapman, L. J., and Pathman, J. H. Errors in the Diagnosis of Mental Deficiency in Schizophrenia - 432

Chorost, S. B., Spivack, G., and Levine, M. Bender-Gestalt Rotations and EEG Abnormalities in Children - 559

Clayton, W. H. See *Ellsworth, R. B.*

Cohen, J. The Factorial Structure of the WISC at Ages 7–6, 10–6, and 13–6 - - - - - 285

Couch, A. S. See *Kassebaum, G. G.*

Clayton, H., and Payne, D. Validation of Doppelt's WAIS Short Form with a Clinical Population - 467

Cowan L., and Goldman, M. The Selection of the Mentally Deficient for Vocational Training and the Effect of this Training on Vocational Success - - - - - - - - - - - 78

Cowen, E. L., and Tongas, P. N. The Social Desirability of Trait Descriptive Terms: Applications to a Self-Concept Inventory - - - - - - - - - - - - - - - - - 361

DeSoto, C. B., and Kuethe, J. L. The Set to Claim Undesirable Symptoms in Personality Inventories 496
Duffendack, S. C. See *Lasky, J. J.*
Edwards, A. L., Wright, C. E., and Lunneborg, C. E. A Note on "Social Desirability as a Variable in the Edwards Personal Preference Schedule" 558
Eichman, W. J. Discrimination of Female Schizophrenics with Configural Analysis of the MMPI Profile 442
Ellen, P. See *Maier, N. R. F.*
Ellis, A. Requisite Conditions for Basic Personality Change 538
Ellsworth, R. B., and Clayton, W. H. Measurement of Improvement in "Mental Illness" 15
Estes, Betsy W., Kodman, F., and Akel, M. The Validity of the Columbia Mental Maturity Scale 561
Fairweather, G. W. See *Lindemann, J. E.*
Fisher, J. The Twisted Pear and the Prediction of Behavior 400
Fisher, S. Body Reactivity Gradients and Figure Drawing Variables 54
Fiske, D. W., and VanBuskirk, C. The Stability of Interpretations of Sentence Completion Tests 177
Friedman, J. Weight Problems and Psychological Factors 524
Garfield, S. L., and Sineps, J. An Appraisal of Taulbee and Sisson's "Configurational Analysis of MMPI Profiles of Psychiatric Groups" 333
Gieseking, C. F. See *Williams, H. L.*
Ginott, H. C. The Theory and Practice of "Therapeutic Intervention" in Child Treatment 160
Glickstein, M. A Note on Wittenborn's Factor Analysis of Rorschach Scoring Categories 69
Goldberg, L. R. The Effectiveness of Clinicians' Judgments: The Diagnosis of Organic Brain Damage from the Bender-Gestalt Test 25
Goldman, M. See *Cowan, L.*
Goodstein, L. D., and Heilbrun, A. B., Jr. The Relationship between Personal and Social Desirability Scale Values of the Edwards Personal Preference Schedule 183
Gordon, J. E., Martin, B., and Lundy, R. M. GSRs during Repression, Suppression, and Verbalization in Psychotherapeutic Interviews 243
Gordon, J. E. Relationships among Mothers' *n* Achievement, Independence Training Attitudes, and Handicapped Children's Performance 207
Gottesman, I. I. More Construct Validation of the Ego-Strength Scale 342
Greenfield, N. S. The Relationship between Recalled Forms of Childhood Discipline and Psychopathology 139
Griffith, A. V., and Peyman, D. A. R. Eye-Ear Emphasis in the DAP as Indicating Ideas of Reference 560
Gross, L. R. Effects of Verbal and Nonverbal Reinforcement in the Rorschach 66
Gross, L. R. MMPI *L-F-K* Relationships with Criteria of Behavioral Disturbance and Social Adjustment in a Schizophrenic Population 319
Hanley, C. Responses to the Wording of Personality Test Items 261
Harway, N. I. Some Factors in Psychotherapists' Perception of their Patients 379
Hays, W. L. A Note on an Index for the Comparison of Profiles on the "Circle" 541
Heilbrun, A. B., Jr. An Immediate Memory Modification of the Block Design Task: Relative Performances of Brain-Damaged and Control Subjects 390
Heilbrun, A. B., Jr. Validation of a Need Scaling Technique for the Adjective Check List 347
Heilbrun, A. B., Jr. See *Goodstein, L. D.*
Heilizer, F. A Note on the Postural Sway Distribution 493
Hiler, E. W. The Sentence Completion Test as a Predictor of Continuation in Psychotherapy 544
Hellmer, L. A. See *Becker, W. C.*

Hover, G. L. See *Lasky, J. J.*
Hovey, H. B., Kooi, K. A., and Thomas, M. H. MMPI Profiles of Epileptics - - - - - 155
Howard, W. Validities of WAIS Short Forms in a Psychiatric Population - - - - - - 282
Hoyt, T. E., and Baron, M. R. Anxiety Indices in Same-Sex Drawings of Psychiatric Patients with High and Low MAS Scores - - - - - - - - - - - - - - - - - 448
Hozier, Ann. On the Breakdown of the Sense of Reality: A Study of Spatial Perception in Schizophrenia - 185
Hunt, D. E. See *Allison, J.*
Izard, C. E. Paranoid Schizophrenic and Normal Subjects' Perceptions of Photographs of Human Faces - 119
Jensen, A. R. A Statistical Note on Racial Differences in the Progressive Matrices - - - 273
Johnson, D. E., Bauer, R. W., and Brown, D. R. Visual Afterphenomena in Diagnosis - - 90
Kagan, J. The Stability of TAT Fantasy and Stimulus Ambiguity - - - - - - - - 266
Kassebaum, G. G., Couch, A. S., and Slater, P. E. The Factorial Dimensions of the MMPI 226
Katzenmeyer, W. G. See *Spielberger, C. D.*
Kausler, D. H. See *Trapp, E. P.*
Klett, C. J., and Lasky, J. J. Agreement among Raters on the Multidimensional Scale for Rating Psychiatric Patients - 281
Kodman, F. See *Estes, Betsy W.*
Kooi, K. A. See *Hovey, H. B.*
Kuethe, J. L. See *DeSoto, C. B.*
Lasky, J. J., Hover, G. L. Smith, P. A., Bostian, D. W., Duffendack, S. C., and Nord, C. L. Post-Hospital Adjustment as Predicted by Psychiatric Patients and by their Staff - 213
Lasky, J. J. See *Klett, C. J.*
Lawton, M. P. The Screening Value of the Cornell Medical Index - - - - - - - - - 352
Lesser, G. S. Population Differences in Construct Validity - - - - - - - - - - - 60
Levin, M. L. A Comparison of the Effects of Phenobarbital, Promethazine, Chlorpromazine, and Placebo Upon Mental Hospital Patients - - - - - - - - - - - - - - - 167
Levin, M. L. A Reply to Sines' Note on a Comparison of the Effects of Phenobarbital, Promethazine, Chlorpromazine, and Placebo upon Mental Hospital Patients - - - 463
Levine, M., Spivack, G., and Wight, B. The Inhibition Process, Rorschach Human Movement Responses, and Intelligence: Some Further Data - - - - - - - - - - - 306
Levine, M. See *Chorost, S. B.*
Levine, M. See *Spivack, G.*
Lighthall, F., Ruebush, B., Sarason, S., and Zweilbelson, I. Change in Mental Ability as a Function of Test Anxiety and Type of Mental Test - - - - - - - - - - - - 34
Lindemann, J. E., Fairweather, G. W., Stone, G. B., Smith, R. S., and London, I. T. The Use of Demographic Characteristics in Predicting Length of Neuropsychiatric Hospital Stay - 85
Liverant, S. MMPI Differences between Parents of Disturbed and Nondisturbed Children 256
London, I. T. See *Lindemann, J. E.*
Lubin, A. See *Rioch, Margaret J.*
Lubin, A. See *Williams, H. L.*
Lundy, R. M. See *Calden, G.*
Lundy, R. M. See *Gordin, J. E.*
Lunneborg, C. E. See *Edwards, A. L.*
MacBrayer, Caroline T. Symbolism Validity and Learning without Awareness - - - - 357
McDavid, J., Jr. Some Relationships between Social Reinforcement and Scholastic Achievement - 151
McGowan, J. F. See *Schmidt, L. D.*
Madoff, J. M. The Attitudes of Mothers of Juvenile Delinquents toward Child Rearing - 518
Maier, N. R. F., and Ellen, P. The Integrative Value of Concepts in Frustration Theory 195

Mann, Carola H. See *Mann, J. H.*

Mann, J. H., and Mann, Carola H. Insight as a Measure of Adjustment in Three Kinds of Group Experience - - - - - - - - - 91

Marlowe, D. Relationships among Direct and Indirect Measures of the Achievement Motive and Overt Behavior - - - - - - - - - 329

Martin, B. The Validity of a Self-Report Measure of Anxiety as a Function of the Time Interval Covered by the Instructions - - - - - - - - - 468

Martin, B. See *Gordon, J. E.*

Masling, J. The Effects of Warm and Cold Interaction on the Administration and Scoring of an Intelligence Test - - - - - - - - - 336

Mednick, S. A. See *Affleck, D. C.*

Mischel, W. See *Pile, E.*

Mitchell, Anne C. A New Maximum CA for the Draw-A-Man Test - - - - - - - - - 555

Moos, R., and Mussen P. Sexual Symbolism, Personality Integration, and Intellectual Functioning - - - - - - - - - 521

Murstein, B. I. Conceptual Model of Projective Techniques Applied to Stimulus Variations with Thematic Techniques - - - - - - - - - 3

Mussen, P. See *Moos, R.*

Nebergall, Nelda S., Angelino, H., and Young, H. H. A Validation Study of the Self-Activity Inventory as a Predictor of Adjustment - - - - - - - - - 21

Nichols, R. C. The Effect of Ego Involvement and Success Experience on Intelligence Test Results - - - - - - - - - 92

Nord, C. L. See *Lasky, J. J.*

Otis, L. S. What Does the Rorschach *Z* Score Reflect? - - - - - - - - - 373

Pathman, J. H. See *Chapman, L. J.*

Payne, D. See *Clayton, H.*

Peterson, D. R., Cattell, R. B., and Quay, H. C. Personality Factors in Nursery School Children as Derived from Teacher Ratings - - - - - - - - - 562

Peterson, D. R., Quay, H. C., and Anderson, A. C. Extending the Construct Validity of a Socialized Scale - - - - - - - - - 182

Peterson, D. R., Quay, H. C., and Cameron, G. R. Personality and Background Factors in Juvenile Delinquency as Inferred from Questionnaire Responses - - - - - - - - - 395

Peterson, D. R. See *Becker, W. C.*

Peyman, D. A. R. See *Griffith, A. V.*

Philbrick, Emily B. The Validity of the Spiral Aftereffect as a Clinical Tool for Diagnosis of Organic Brain Pathology - - - - - - - - - 39

Pile, E., Mischel, W., and Bernstein, L. A Note on Remoteness of TAT Figures as an Interpretive Concept - - - - - - - - - 252

Psychological Test Reviews - - - - - - - - - 283, 470, 563

Quay, H. C. See *Becker, W. C.*

Quay, H. C. See *Peterson, D. R.*

Rabin, A. I. A Contribution to the "Meaning" of Rorschach's Inkblots via the Semantic Differential - - - - - - - - - 368

Richey, Marjorie H., and Spotts, J. V. The Relationship of Popularity to Performance on the Goodenough Draw-A-Man Test - - - - - - - - - 147

Rioch, Margaret J., and Lubin, A. Prognosis of Social Adjustment for Mental Hospital Patients under Psychotherapy - - - - - - - - - 313

Roen, S. R. Correlates of Attitudes among Clinical Psychology Students - - - - - - - - - 533

Rozynko, V. V. Social Desirability in the Sentence Completion Test - - - - - - - - - 280

Ruebush, B. See *Lighthall, F.*

Rumrill, C. See *Wiggins, J. S.*

Rychlak, J. F. Forced Associations, Symbolism, and Rorschach Constructs - - - - - - - - - 455

Sarason, *S.* See *Lighthall*, *F.*
Saucer, *R. T.* Chlorpromazine and Apparent Motion Perception by Schizophrenics - - - 134
Schein, *J. D.* Note on a Bibliographical Resource for Studies of Afterimages - - - - - 137
Schmidt, *L. D.*, *McGowan*, *J. F.* The Differentiation of Human Figure Drawings - - - 129
Schonbar, *Rosalea A.* Some Manifest Characteristics of Recallers and Nonrecallers of Dreams - 414
Schubert, *D. S. P.* Personality Implications of Cigarette Smoking among College Students 376
Schlafer, *R. J.* See *Calden*, *G.*
Shippee-Blum, *Eva Maria.* The Young Rebel: Self-Regard and Ego-Ideal - - - - - - 44
Shoemaker, *D. J.* See *Becker*, *W. C.*
Sines, *L. K.* A Note on Levin's Report of Findings with Phenobarbital, Promethazine, Chlorpromazine, and Placebo - 461
Sines, *L. K.* The Relative Contribution of Four Kinds of Data to Accuracy in Personality Assessment - 483
Sineps, *J.* See *Garfield*, *S. L.*
Singh, *T. R.* See *Sinha*, *D.*
Sinha, *D.*, *and Singh*, *T. R.* Manifest Anxiety and Performance on Problem Solving Tasks 469
Slater, *P. E.* See *Kassebaum*, *G.*
Smith, *E. E.* Defensiveness, Insight, and the *K* Scale - - - - - - - - - - - - - 275
Smith, *R. S.* See *Lindemann*, *J. E.*
Smith, *P. A.* See *Lasky*, *J. J.*
Speisman, *J. C.* Depth of Interpretation and Verbal Resistance in Psychotherapy - - - 93
Sperrazzo, *G.*, *and Wilkins*, *W. L.* Racial Differences on Progressive Matrices - - - - 273
Spielberger, *C. D.*, *and Katzenmeyer*, *W. G.* Manifest Anxiety, Intelligence, and College Grades - 278
Spivack, *G.* See *Chorost*, *S. B.*
Spivack, *G.* See *Levine*, *M.*
Spotts, *J. V.* See *Richey*, *Marjorie H.*
Spivack, *G.*, *Levine*, *M.*, *and Sprigle*, *H.* Intelligence Test Performance and the Delay Function of the Ego - 428
Sprigle, *H.* See *Spivack*, *G.*
Stieper, *D. R.*, *and Wiener*, *D. N.* The Problem of Interminability in Outpatient Psychotherapy - 237
Stone, *Beth.* Prognostic Factors in Schizophrenia - - - - - - - - - - - - - - 279
Stone, *G. B.* See *Lfndemann*, *J. E.*
Streitfeld, *J. W.* Expressed Acceptance of Self and Others by Psychotherapists - - - - 435
Stumpf, *J.* See *Beier*, *E. G.*
Taylor, *J. B.* Social Desirability and MMPI Performance: The Individual Case - - - - 514
Thomas, *M. H.* See *Hovey*, *H. B.*
Tongas, *P. N.* See *Cowen*, *E. L.*
Trapp, *E. P.*, *and Kausler*, *D. H.* Association Tendencies of Groups Differentiated on the Taylor Manifest Anxiety Scale - - - - - - - - - - - - - - - - - - 387
Van Buskirk, *C.* See *Fiske*, *D. W.*
Wiener, *D. N.* See *Stieper*, *D. R.*
Wiener, *G.* The Interaction among Anxiety, Stress Instructions, and Difficulty - - - - 324
Wiggins, *J. S.* Interrelationships among MMPI Measures of Dissimulation under Standard and Social Desirability Instructions - - - - - - - - - - - - - - - - - 419
Wiggins, *J. S.*, *and Rumrill*, *C.* Social Desirability in the MMPI and Welsh's Factor Scales *A* and *R* - 100
Wight, *B.* See *Levine*, *M.*
Wilkins, *W. L.* See *Sperrazzo*, *G.*

Williams, H. L., Lubin, A., and Gieseking, C. F. Direct Measurement of Cognitive Deficit in Brain-Injured Patients - 300
Wishner, J. Factor Analysis of Rorschach Scoring Categories and First Response Times in Normals - 406
Wittenborn, J. R. Some Comments on Confounded Correlations Among Rorschach Scores 75
Wohl, J. See *Capell, M. D.*
Wright, C. E. See *Edwards, A. L.*
Young, H. H. See *Nebergall, Nelda S.*
Zuk, G. H. Autistic Distortions in Parents of Retarded Children - - - - - - - - - - 171
Zweilbelson, I. See *Lighthall, F.*

JOURNAL OF CONSULTING PSYCHOLOGY

February, 1959 Vol. 23, No. 1

Contents

EDITORIAL: *Edward S. Bordin* - - - - - 1

ARTICLES

A Conceptual Model of Projective Techniques Applied to Stimulus Variations With Thematic Techniques: *Bernard I. Murstein* - - - - - 3

Measurement of Improvement in "Mental Illness": *Robert B. Ellsworth and William H. Clayton* - - - - - 15

A Validation Study of the Self-Activity Inventory as a Predictor of Adjustment: *Nelda S. Nebergall, Henry Angelino, and Harl H. Young* - - - - - 21

The Effectiveness of Clinicians' Judgments: The Diagnosis of Organic Brain Damage From the Bender-Gestalt Test: *Lewis R. Goldberg* - - - - - 25

Change in Mental Ability as a Function of Test Anxiety and Type of Mental Test: *Frederick Lighthall, Britton Ruebush, Seymour Sarason, and Irving Zweibelson* - - - 34

The Validity of the Spiral Aftereffect as a Clinical Tool for Diagnosis of Organic Brain Pathology: *Emily B. Philbrick* - - - - - 39

The Young Rebel: Self-Regard and Ego-Ideal: *Eva Maria Shippee-Blum* - - - - - 44

An Approach to the Factor Structure of Clinical Judgments: *Martin D. Capell and Julian Wohl* - - - - - 51

Body Reactivity Gradients and Figure Drawing Variables: *Seymour Fisher* - - - - - 54

Population Differences in Construct Validity: *Gerald S. Lesser* - - - - - 60

Effects of Verbal and Nonverbal Reinforcement in the Rorschach: *Leonard R. Gross* - 66

A Note on Wittenborn's Factor Analysis of Rorschach Scoring Categories: *Mitchell Glickstein* - - - - - 69

Some Comments on Confounded Correlations Among Rorschach Scores: *J. R. Wittenborn* 75

The Selection of the Mentally Deficient for Vocational Training and the Effect of This Training on Vocational Success: *Lawrence Cowan and Morton Goldman* - - - - - 78

The Use of Demographic Characteristics in Predicting Length of Neuropsychiatric Hospital Stay: *James E. Lindemann, George W. Fairweather, Gideon B. Stone, Robert S. Smith, and Ira T. London* - - - - - 85

BRIEF REPORTS

Visual Afterphenomena in Diagnosis: *Derwood E. Johnson, Robert W. Bauer, and Don R. Brown* - - - - - 90

Insight as a Measure of Adjustment in Three Kinds of Group Experience: *John H. Mann and Carola Honroth Mann* - - - - - 91

The Effect of Ego Involvement and Success Experience on Intelligence Test Results: *Robert C. Nichols* - - - - - 92

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 80 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual.* Headings, tables, and references are avoided or, if essential, must be counted in the 80 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 80-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual.*

EDITORIAL

With this issue, the *Journal of Consulting Psychology* has become the responsibility of a new Editor, punctuating 12 years of guidance by Laurance F. Shaffer. I do not plan to mark this occasion by any extended review of that significant period nor by an elaborate statement of a prospectus. The retiring Editor plans a report of his stewardship, and to him naturally falls the responsibilities and well earned satisfactions such a review will offer.

An editor of a well established journal, as this one is, finds little scope for radical invention. The American Psychological Association, through its Publications Board, defines its general sphere as "the clinical journal." As such, its contents include original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Although quantitative studies have been given priority, relevant theoretical contributions, case studies, and descriptions of clinical techniques are also acceptable.

Since readers and authors are a vastly overlapping group, I am using this opportunity of catching their attention to emphasize the difference between the official definition of the Journal's scope and what is defined by the articles that appear between its covers. The primacy of original research is easily evident even to the casual reader. The retiring Editor, with the help of other research minded clinical psychologists, has removed any lasting doubt that the subject matter of clinical psychology is susceptible to thoroughgoing empirical treatment. From here on, the image of the journal must be sketched in largely imaginary lines. Theoretical contributions occur at the rate of slightly more than one a year. Case studies and descriptions of clinical techniques are virtually unrepresented.

Why the discrepancy between the official and the functioning images? Some may suggest that it reflects the Editor's biases. Experience with my first 200 submitted manuscripts makes me doubt the validity of that explanation. In that sample, there were the usual one or two theoretical contributions that were worthy enough to be accepted for publication. The two case studies offered were done without clear focus as to their purpose or with conviction that they could be scholarly contributions.

I refuse to believe that the potential for a more rounded contribution to the literature on clinical psychology does not exist. I affirm my faith that we will not endanger our newly won rigor and empiricism when we give our attention and the Journal's pages to imaginative flights beyond the fully charted regions, to speculative excavations that will open up new mines of data, and to the vivid presentation of clinical experiences from which research in clinical psychology, above all else, should draw its nourishment.

There ought to be more evidence of the ferment in our field. We need reviews that force all of us to pause in our mad chase for data to contemplate what the data at hand have to offer. When so many studies of manifest anxiety have been completed, it is time for someone to examine the question of what we have learned and of which are the most profitable new directions to pursue. Incidentally, too many manuscripts hurry to tell what data have been collected, how obtained, and the results of their analysis—before taking any trouble to show the reader why he should care what answer is forthcoming. I have heard the opinion expressed that the Journal wants short papers; authors are expected to excise all theory. This is not to be our policy. Similarly, even this brief experience has made me acutely conscious of how much fragmenting of research reports exists. Within one volume, or distributed over two, a man will publish two or more interrelated studies that would best be reported as one. I shall act on the conviction that one well rounded report incorporating a series of related studies is worth more than the sum of the piecemeal reports of the parts of the series.

As I write, I find that a fighting mood rises. I do think we need some controversy, but without rancor. In constructive fashion, we need to review each other's work. We sorely need critical reviews of specific tests or clinical techniques, or of ways of reaching specific types of inferences, of interest to the clinician. The data regarding many of our most used clinical instruments, especially the Rorschach and Minnesota Multiphasic, are so vast and so heterogeneous that they defy review as a whole. It seems to me that there can be no meaningful answer to the question of whether either instrument is valid. It is more meaningful to take a particular type of inference that is drawn by use of the instrument, examine the logical structure on which it rests, and review the data which will show how tenable that logical structure is.

Finally, there is a need for communication and mutual stimulation that can lead to the evolution of new clinical procedures. We need discussions and descriptions on such topics as patterns of group practice, intake practices, or problems of supervision—to mention but a few.

Some may fear that I am inviting rambling, indistinctly formed, free associative contributions such as might come from a diary of a clinical psychologist. This is far from my purpose. The *Journal of Consulting Psychology* will accept only those contributions that show evidence that the creative spark has been fanned into flame through the application of a rigorous scholarship which places the contribution in its theoretical and historical context.

EDWARD S. BORDIN
Editor

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

A CONCEPTUAL MODEL OF PROJECTIVE TECHNIQUES APPLIED TO STIMULUS VARIATIONS WITH THEMATIC TECHNIQUES

BERNARD I. MURSTEIN
University of Portland

The clinical psychologist today would hardly accept without reservation the view that the projective techniques necessarily evoke responses tapping the ". . . private world of the individual, comprising as it does, the feelings, urges, beliefs, attitudes, and desires of which he may be dimly aware and which he is often reluctant to admit even to himself, much less to others . . ." (Frank, 1948, p. 66). Neither would he, in most cases, leap to second the view that "The interpretative methods [TAT] offer the subject a situation or action to which he responds by a creative activity, wherein is disclosed his basic concepts, expectations, feelings" (Frank, 1948, p. 57).

Such views have not sufficiently emphasized the importance of the stimulus properties of the various techniques as well as such background characteristics as examiner-examinee relationship in determining the response elicited. What is currently needed is a theory to account for the interaction of the various determinants of responses to projective techniques. In this regard, the theory of adaptation-level of Helson (1955) seems most helpful. The theory briefly stated would hold that:

> Operationally the adaptation level is represented by the stimulus to which the organism responds, either not at all or in an indifferent or neutral manner. In a large variety of situations the AL proves to be a weighted log mean of three classes of stimuli: the stimulus in the focus of attention; all stimuli in the field forming the context or background; and residuals from past experience. These three classes of stimuli pool or interact to determine the AL and, hence, the adjustment of the organism. As stimulation or behavior varies, the adaptation level fluctuates accordingly. A simple formula expresses these facts more clearly and concisely as follows:
>
> $$\log A = p \log X + q \log B + r \log R$$
>
> where p, q, and r are weighting factors showing the relative importance of X—the stimulus, B—the background, and R—the residual (pp. 91–92).

In adapting this approach to the analysis of responses to projective techniques, it may be useful to spell out each of the three categories in some detail.

Stimulus Properties of Tests

Under this heading may be considered the varying degrees of ambiguity characteristic of the different techniques. Thus, the Draw-A-Person has no structure aside from the limitations ensuing from the instructions to "draw a whole person," and the use of an 8 × 11 sheet of white paper and a pencil. The Rorschach has a series of blots selected on the basis of certain shading, color, and structural qualities. The TAT, which is more heterogeneous in its make-up, contains some cards in which the characters are difficult to distinguish and other cards in which the characters involved are easily distinguished, but the feelings they are experiencing or actions engaged in are not readily apparent.

Background Characteristics

Here one deals with the environment both psychological and physical in which the test is given. Questions of concern are the purpose of the testing from the point of view of the subject and from the point of view of the examiner. Is the test given to obtain normative data from college sophomores, or to determine whether a "model prisoner" is to be paroled? Other variables influencing the response will be the examiner's sex, age, attractiveness, and interest in the subject.

Personality Characteristics

Within the construct of personality one deals with the subcategories: (*a*) organismic needs and (*b*) learned needs. Examples of organismic needs are hunger, sex drive, thirst, and fatigue. Learned needs have to do with cultural values and ego needs. The two, of course, may well interact in that one might drink a cup of tea (organismic need) while holding the cup and saucer in a culturally prescribed manner (learned need).

The perception of such needs on a projective technique is a function (*a*) of the intensity of the need and (*b*) the expectancy of satisfying that need. These factors have been pointed out by Rotter (1954). The importance of Factor *b* in the manifestation of the projected need may be illustrated through reference to the work of Brozek, Guetzkow, and Baldwin (1951). Using males subjected to a semistarvation diet for 24 weeks, the authors reported no significant increase in perception of food on the Rorschach. It should be apparent, therefore, that *the presence of a need without the means of satiating it may inhibit the perception of objects relating to the need rather than sensitizing one to them*. The work of Levine, Chein, and Murphy (1942) also seems to substantiate this conclusion.

Lastly, the *S* himself interprets the stimulus, background and variables applying to the self and, either consciously or unconsciously, arrives at a decision as to what he will report to the examiner. The end response may thus be viewed as composed of two components, that which is within the control of the perceiver and that which is beyond his control (i.e., where the needs are overt and too strong to control, or are unacceptable and covert as in unconscious needs).

It would seem, then, that in addition to underestimating the importance of the stimulus and background factors, many clinicians have underestimated the strength of the control which many *S*s exercise over the responses they manifest. It has been assumed that the *S* must reveal his "private world." If he does so consciously there is no difficulty save that the same information might have been obtained more economically (both from the point of view of time as well as money) through a brief interview. If, on the other hand, the *S*'s defenses prevent the free expression of his problem, it has been assumed that the character of his defenses may be analyzed in order to get at the problem. In short, no response is held to be an "accident." In its broadest sense this statement is doubtlessly true. From an empirical standpoint, however, many responses are "error variance" or superfluous to the problem which the clinician seeks to answer. Thus, on the Rorschach one might obtain a response to the white space of Card 2 of "A new type of jet plane. I saw it yesterday on television." The response ostensibly may have little to add to the understanding of the *S*. But the clinician might think, "Why of all the possible objects he saw yesterday, did he happen to report a jet airplane?" The *S* might well state that he had seen other planes, ships, etc., but that this particular space seemed to rather closely resemble "the delta-swept wing jet." Such an answer would rarely satisfy some clinicians. It is unfortunate that the effect of stimulus properties in determining a response is so thoroughly neglected in favor of wholly subject-oriented interpretations.

Another obstacle in the understanding of projective techniques has been the failure to accept the view that even with severely disturbed persons, some defenses may be impossible to pierce from an analysis of the data. The omission of response as occurs in a Rorschach record totaling only five responses, is probably indicative of "something," but that "something" will rarely be apparent from the analysis of the five responses.

If we are to build a theory of projective techniques, much more data are needed with regard to the roles played by the stimulus, background, and personality variables. This paper undertakes an analysis of the effect of stimulus variation of TAT-type pictures toward the end of systemizing and interpreting some of the data already obtained. It will attempt to list the key studies in this area and to clarify the importance of the findings for projective test theory.

Stimulus Variations with TAT-Type Tests

Stimulus variations have been used with regard to light control, length of time cards have been exposed, changes in the construction of the drawings, and number of possible interpretations that might apply to each card. It is proposed to consider the work done in each of these areas to determine their effect upon the responses obtained.

Variations in Lighting

Weisskopf (1950a) had college students write descriptions of several TAT cards which were presented either in an unaltered form, or as pictures taken under reduced exposure, giving the pictures a hazy effect. No significant differences in "transcendence" (mean number of responses going beyond pure description) were obtained. More recently, Bradley and Lysaker (1957), using a lifelike TAT type picture, varied the illumination from normal to three successive darker stages of illumination, as well as three lighter ones. The *Ss* were several hundred housewives (range 122 to 171 for each stage) with each *S* seeing the picture in only one stage. Similar to Weisskopf's (1950a) finding, no difference in productivity of response was noted. In analyzing the content of associations with regard to the picture, however, a clear positive linear relationship was found between increasing degree of darkness and pleasantness of association. When the picture was used with a varying degree of increasing light, the picture slightly lighter than the control elicited the most favorable association, followed by the next lighter picture, the lightest picture, and the control picture. The effect of light was apparently not a linear function, with a moderate amount of supranormal illumination considered most pleasant, and either more or less illumination considered less pleasant.

It seems clear that with the absence of light, the stimulus properties of the picture became increasingly vague, thus resulting in an increasing "internal" perception on the part of the *S*. In the relative absence of sensory stimulation, a person finds it easier to introspect and to relate personal material with considerably less inhibition than under normal circumstances. The soothing yet fantasy enriching effect of darkness is what makes a motion picture more enjoyable in a "movie" house than under the semilighted atmosphere of a television room. Moreover, many psychotherapists have their seating arranged so as to be out of the patients' direct gaze on the assumption that there is less interference with the patients' private thoughts. Here, too, the darkened room is often more likely to educe free association than is the normally lighted one.

The supra-illumination effects are more difficult to interpret. Possibly, a limited amount of extra light obscures the picture and permits pleasant fantasies to be evoked. An increasing amount of light may be irritating to the nervous system and accordingly elicit less pleasant associations. The relative lack of "pleasantness" occurring under normal illumination may have stemmed from the stimulus properties being most clear at this stage, and hence the subjects having been more likely to respect the neutral emotional tone of the picture (woman baking a cake).

The question may arise as to why an association elicited under less than optimal sensory conditions should be pleasant rather than unpleasant. Two factors may be suspected of playing a part here. In the first place, fantasy often occurs as a temporary vicarious means of satisfying needs and tensions which are as yet unsatisfied in the outer world. The teen-age girl dreams of a prince charming, the young executive imagines himself as a junior vice-president. Many of our fantasies are thus related to a need to succeed in a variety of activities, the attainment of which would be associated with pleasant feelings. It is, of course, true that unpleasant, even hostile, tensions may also be present. The young girl may be annoyed because a rival seems to have currently won her prince charming; the young executive may chafe under the imposition of being a "yes" man to the "boss." Nevertheless, militating against the appearance of hostility on the TAT would be the fact that the *S* would probably deem the testing situation as inappropriate for the expression of hostility. The inhibitory effect of the "background" might be due to a weak generalization gradient from hostility to the "boss" to hostility in an ambiguous situation, or the *S* might simply inhibit all expression of hostility because he felt he was being judged and he wanted to make a good impression. Since it is socially acceptable to perceive pleasant things and not as acceptable to perceive hostility in others, it is more than likely that when the stimulus properties of the object being perceived are weak enough, the *S* will be apt to fantasize publicly only over a narrow band of ideation, namely, those topics which do not conflict with what the *S* perceives to be the ego demands of the social situation in which he is involved.

Variations in Time Exposure

Weisskopf (1950a) presented college *S*s with 2 sets of TAT pictures exposed for .2 of a second and for 5 seconds. The latter group manifested a higher number of fantasy scores ("transcendence"). Kenny (1954) exposed the TAT pictures to college *S*s for 5 seconds and 2 minutes, respectively. The correlation between "personality revealingness" of the themes elicited and "transcendence" was .62 and .64 for the 5-second and 2-minute conditions of exposure.

Apparently, the individual fantasizes with greater personal involvement as well as greater productivity when he has sufficient time to appraise the stimulus qualities of the card. After a sufficient time has elapsed (optimally, perhaps, a few seconds), added length of exposure has little effect.

Variations in the "Ground"

Completely traced line drawings proved to be more effective in evincing fantasy than incompletely traced ones, in two studies by Weisskopf (1950b) and Weisskopf-Joelson and Lynn (1953). Bradley and Lysaker (1957) reported that varying the background of a picture, when the figure (woman baking a cake) remained constant, had no effect insofar as altering the productivity of responses to the pictures. These studies seem to indicate that ambiguity of the figure affects fantasy production, but alteration of the background has little effect provided the figure remains clear.

Variations of Central Figures

Thompson Modification. Murray (1943) suggested that at least one card should be chosen, showing a figure of approximately the same age and sex as the *S*. Tomkins (1947), on the other hand, has proposed that the TAT may be interpreted most meaningfully by taking into account the psychological distance of the stimulus from the *S*. Moreover, the meaningfulness of the responses elicited by the TAT should be a function of the "remoteness" of the thematic material presented.

Thompson (1949), adhering to Murray's suggestion, believes that the closer the stimulus resembles the actual *S*, the more the *S* will identify with the figure and, accordingly, be likely to produce more meaningful material. To test this hypothesis, he constructed a set of TAT cards similar to the original TAT, except that Negro characters were substituted for white ones. Using 26 Negro college students from a Negro school, the Thompson TAT was group-administered by a Negro examiner who flashed the cards on a screen by means of a projector. The *S*s were instructed to write stories to the projected pictures. In order to compare the Murray and Thompson versions, the *S*s were divided into two groups of 13 each, with half receiving the Thompson series first and the Murray version second, while the other half received both series with the order of presentation reversed. Thompson found a significant increase ($p < .01$) in story length to the Thompson TAT for each of the 10 cards used.

More recent studies, however, have been highly critical of the value of the Thompson modification. Riess, Schwartz, and Cottingham (1950) used 30 Negroes and 30 white females from Hunter College in New York City. The usual AB, BA order for the two TAT tests was followed, using a Negro examiner for half of both the Negro and white group, and a white examiner for the other half. The results were:

> Negroes and whites in the North produce stories that differ insignificantly in length regardless of whether the stimulus material is Negro or not, and regardless of the color of the examiner, with the exception of a tendency for northern whites to increase

story length on Negro stimulus material with a Negro examiner (Riess et al., 1950, p. 708).

In another article using the same data (Schwartz, Riess, and Cottingham, 1951), the authors investigated the number of ideas appearing in the stories rather than merely the story length. Their data showed that (*a*) Negroes expressed the most ideas when given the cards by a white examiner. The set of cards used was of little consequence. (*b*) Whites gave more ideas when the Thompson set was used. This result was not a function of whether the examiner was white or Negro. (*c*) The production of ideas was quite low when the stimulus, examiner, and *S*s were of the same race.

Korchin, Mitchell, and Meltzoff (1950) used two groups of 80 Negro and 80 white male *S*s from Philadelphia, half of each group being "middle class" while the other half was of "low" socioeconomic status. The examiner was white. An analysis of variance showed no significant effects due to race, nor any significant interaction between race and status for story length. Only the "class" differences were significant ($p < .01$), with the "middle class" telling the longer stories.

In still another experiment, Light (1955) divided 26 white students into two groups of 13 each, one-half receiving the Murray TAT, while the other received the Thompson version. No significant differences in story length were found, although Light found certain themes to be more frequent with the Thompson set. These themes were crime, poverty, occupational inferiority, witchcraft, and prostitution.

Cook (1953) used 60 male college students (30 Negro, 30 white), divided into 4 groups of 15 each. One-half of the Negro and white groups each received the Thompson TAT, while the remaining two groups receiving the Murray TAT were composed of the remaining Negro and white *S*s, respectively. The examiner here also was white, and most of the *S*s had spent at least 15 years in the South. Cook wished to examine the relationship between tendency toward ego-defensiveness as a function of a decrease in remoteness between the *S* and the stimuli used. The measures of ego-defensiveness used were word count, compliance with instructions given, vagueness of stories, number of words indicating uncertainty, number of alternatives offered, number of references to the pictures (i.e., excessive use of description), and number of different themes offered. No significant differences solely attributable to the two sets of pictures were found. Both the Negroes and whites were interviewed with regard to their perception of the test. It was found that the great majority of the Negroes perceived both the Murray and Thompson TATs as dealing with "people in general." On the other hand, the whites perceived the Thompson TAT as dealing with Negroes rather than "people in general." Hence, it would be expected that the chief differences between the Negro and white *S*s would be found with regard to the Thompson TAT. This was precisely what happened. The Murray TAT yielded only two significant findings at the .05 level. Negroes used more words indicating uncertainty and made more references to the pictures.

On the Thompson TAT, however, the protocols of the whites had a significantly higher word count. The Negroes had a greater number of alternatives, uncertainty, vagueness, and references to the pictures. All measures were significant beyond the .01 level. Cook believed his findings indicated that the more remote the relationship between the stimulus and the *S*, the less the ego-defensiveness.

The foregoing studies lead to a consideration of the following:

1. Are Negroes to be treated as a homogeneous class? The data hardly suggest that such a treatment is warranted. Apparently, it would be more justifiable if anything to use a "socioeconomic class" approach to interpretation.

2. The degree of similarity between stimulus and subject is of itself insufficient as an explanation of the type of response elicited because it does not take cognizance of the background characteristics. Thus, in Schwartz's (1951) work, Negro *S*s who were administered Negro cards by a Negro examiner manifested a smaller number of ideas in their stories than any other grouping. Negro *S*s receiving the same cards from a white examiner expressed the greatest number of ideas. For the white *S*s, however, the Negro cards

seemed to educe the greatest number of ideas regardless of the race of the examiner.

3. In view of the fact that the Stimulus × Background × Person interaction seems to be of considerable importance in determining the nature of the response, the *t* test favored by previous investigators would appear to be a relatively insensitive statistic. The analysis of variance would appear to be more powerful in this regard.

4. The question arises as to whether story length or number of ideas expressed is an adequate measure of the *S*'s involvement in the TAT. Is it not conceivable that sophisticated *S*s (college students) might manifest an excessive superficial verboseness to mask the projection of manifest or covert needs? What is needed is a quantitative scale measuring the personality meaningfulness of the stories rather than indirect and unproven correlates such as "word-count" and "number of ideas."

5. The influence of the stimulus is not a simple function of its similarity to the *S*. The culture plays a crucial role in the interpretation of perception. Negroes tended to perceive the Murray TAT characters not as "whites" but as "people in general." Whites, however, perceived the Thompson cards as dealing with "Negroes." Thus, majority populations are far more apt to perceive minorities as "different" than vice versa. They also are less apt to have had contact with them. Minority groups such as Negroes, however, must constantly live in a "white" world. Although certainly aware of the special privileges whites possess which they do not, they must adapt perceptually to the majority white culture if they are to be maximally adjusted to their environment. Thus, the finding that Negro girls preferred white dolls to Negro ones (Goodman, 1952) should come as no surprise. When, however, Negroes are confronted with the Thompson figures, particularly in the presence of a white examiner, it is apparent to them that something unusual and different is involved, and they tend to be cautious and vague.

Other Modifications. Lasaga y Travieso and Martinez-Arango (1946), in working with nuns, found no improvement in the diagnostic value of the TAT stories when they substituted nuns for the usual TAT central figures. Weisskopf and Dunlevy (1952) used three groups of 10 each of normal, crippled, and obese male undergraduates. Each *S* received three sets of TAT cards, one of which was unmodified, the other sets being modified so that the central figure was obese in one set and crippled in the other. The *S*s were asked to describe the cards rather than to tell a story. The results indicated that the mean transcendental index (number of elements in the stories going beyond pure description) was significantly less for the obese figures than for the normal and crippled ones. This trend was consistent among all three groups.

In another study, Weisskopf-Joelson and Money (1953) modified the TAT with somewhat crude approximations of the original drawings. In the first series, the head of the central character was a more or less neutral face. In the second session, the same cards had actual photos of the *S*s' heads instead of the original neutral heads. A control group was used to account for the effect of repetition. No significant increase in projection was found for the "photo" set over the neutral set as measured by the transcendence index and word content. In addition, the diagnostic value of the "photo" set when judged by two judges was not found to be any higher than the other set.

In a recent study by Edgar and Shneidman (1958), a group composed of psychotics, psychoneurotics, and persons suffering from personality disorders was used. Material obtained from "patient government" discussions was compared with that obtained from the MAPS technique, and a variation of this technique, in which full length cut-out photos of every person in residence on the ward at the time as well as all staff members, was used. Using Bales Interaction Process Method to analyze the data, the authors found as much reluctance to show antagonism to photos of peer figures as to peers in face-to-face contact, such as occasioned by the group meetings. They suggest that one use of the "photo" technique may be in measuring the feelings of need for evaluation from others and feelings of need for direction. Nevertheless, the authors concluded that "photos, in general, seem to elicit more positive feeling than fantasy figures did, the latter calling forth more overt aggression" (Edgar and Shneidman, 1958, p. 11).

It would appear then that similarity of the central character to the *S* is not only noneffectual in furthering projection but actually may be detrimental. Since obesity is the butt of jokes in our society, it is unlikely that fat persons derive any satisfying tension reduction in identifying with obese characters. Before concluding, however, that similarity per se inhibits projection, it may be of value to recall that the background characteristics (purpose of the experiment) are much more readily apparent to the *S* under these conditions. Thus, while the *S* taking the TAT under the usual nomothetic experimental procedure may not suspect that his responses as an individual will be noteworthy, the *S* who, as in Weisskopf-Joelson and Lynn's study (1953), sees a photo of his head planted somewhat out of context on a crude sketch may not suffer this delusion.

It may be that under nonsocial conditions similarity might well stimulate projection. Indeed, the mirror probably does richly stimulate fantasies in teen-agers as they carefully prune themselves before going out on a "heavy date." Some support for this statement comes from the data of Beier, Izard, Smock, and Tougas (1957) who gave young adults sets of photos of younger, peer, and older persons having them signify the ones liked and disliked. Male *S*s most frequently "liked" peer males while female *S*s preferred peer females.

In sum, however, as far as the usual testing procedure is concerned, it seems clear that similarity to the point of idiosyncratic identification promotes ego defensiveness and a reduction in the degree of projection.

The Use of Animals Instead of Humans

Bellak and Bellak (1949), believing that children identify more readily with pictures containing animal figures than with human ones, created the Children's Apperception Test. Bills (1950) compared a series of pictures of rabbits with several TAT cards using as *S*s children (both male and female) aged five through ten. Significantly longer stories were obtained with the "rabbit" pictures. In another study (Bills, Leiman, & Thomas, 1950), a new group of third-grade *S*s had six play-therapy sessions and then took the "rabbit" pictures test and TAT. Both tests correlated only slightly with the material obtained from play therapy using Murray's rating approach for manifest needs. Their correlations with each other also were rather low, ranging from − 09 to .58.

Biersdorf and Marcuse (1953) used a broad approach to the problem employing CAT cards and cards specially constructed, which were almost identical to the CAT cards except that humans were substituted for the animals while keeping the same emotional expressions. The *S*s were 30 first graders, the group containing both sexes. Not only were no significant differences found between numbers of words used in both tests, but further analysis revealed no difference in length of time before response, length of response time, number of words used, number of ideas used, number of characters mentioned in the pictures, and the number mentioned who were not in the pictures. Mainord and Marcuse (1954), using the same sets of cards, tested emotionally disturbed boys and girls (mean age, seven years) and found similar nonsignificant differences for the aforementioned quantitative measures. They, however, also asked five clinicians to rate the protocols as to "clinical usefulness," the judges being ignorant of the hypotheses of the authors and the stimulus properties of the cards eliciting the protocols. The human figure cards were found to be more "clinically useful" at better than an .001 level of confidence.

In yet another study comparing the CAT with its "humanized" analogue, Furuya (1957) studied the protocols of 72 children ranging in age from 6 to 12. He was interested in answers to the following two questions.

1. With different criteria and age-groups—from those of Biersdorf and Marcuse (1953) and Mainord and Marcuse (1954)—what kind of difference will be found between the productivity to animal pictures to human pictures, equivalent in scenes and situations?
2. Is there any tendency toward relative decrease of productivity to animal pictures

compared with human pictures as the children grow older (Furuya, 1957, p. 248)?

The results are confounded with methods of presentation in that the youngest children received the cards individually, while the older ones received the cards as a group. Nevertheless, the results are consistent with earlier studies in indicating the superiority of the human pictures. Of the 20 *t* tests, four proved to be significant at the .05 level or better. These indicated that the percentage of stories containing expressions of feeling, significant conflict, and having a definite outcome, was significantly higher for the human pictures as compared to the animal ones.

No tendency was found for any increase or decrease in the relative productivity of the animal or human pictures from the youngest to the oldest age range.

Lastly, the author implies that the results might have been even more extensive but for the fact that "most of the animal pictures that have been used, were 'personified or human-like' animal pictures" (Furuya, p. 252).

Light (1954), working with 9- and 10-year-olds, found significant differences between the CAT and TAT. The TAT was found to elicit more feelings, different kinds of feelings, conflicts, number and kinds of outcomes, number and kinds of themes, and number of figures. Only the number of words was not significantly different from test to test. Light believed that this result may have been due to the imposed time limit of seven minutes to write the stories after the pictures had been flashed on a screen.

Armstrong (1954) working with intellectually superior first, second, and third graders also found no difference in number of words used between the CAT and a set of pictures corresponding in style and composition but containing humans instead of animals. The Weisskopf transcendence index, however, indicated that the children gave significantly more nondescriptive statements to the human series.

The results of the foregoing studies do not support the Bellaks' assumption that the use of animal pictures results in an increased facilitation of projection for children. For the most part, where the human and nonhuman drawings have been strictly comparable (Armstrong, 1954; Biersdorf and Marcuse, 1953), the human pictures have been shown to be equal or superior to the animal pictures. The only negative note sounded (Bills, 1950) is not strictly comparable since the "rabbit pictures" utilized may have been conducive to projection *due to the nature of the situation portrayed* on the cards rather than to the presence of animal figures of themselves.

Variations in Ambiguity Between TAT Cards

Murray (1943), in the manual for the Thematic Apperception Test, described the cards as ". . . divided into two series of ten pictures each, the pictures of the second series (numbers 11 through 20), being purposely more unusual, dramatic, and bizarre than those of the first. One full hour is devoted to a series, the two sessions being separated by a day or more" (p. 2).

Weisskopf (1950a) found that college students, when asked to describe both sets, projected significantly more fantasy material to the first "everyday" series than to the latter "fairy tale" one. Of interest also was the fact that women tended to have significantly higher "transcendence" indices than men for both the "male" ($p < .02$) and "female" series ($p < .05$). A heterosexual group of college students were asked by Bijou and Kenny (1951) to rank 21 "male" series TAT cards according to their ambiguity (estimated number of interpretations that might be derived from each card). The authors computed a *t* test between the mean ambiguity level for the first Murray series as compared to the second one. The second "fairy tale" series proved to be more ambiguous, at the .05 level of confidence, although the authors did not reject the null hypothesis in postadopting an .01 level. They also compared the "personality revealingness" value of the first series as compared to the second one by analyzing the themes of the TAT stories (Kenney and Bijou, 1953). The mean "personality revealingness" value of the "everyday" series proved to be somewhat higher than the "fairy tale" series, but the *t* test of .92 was not statistically significant.

Using the same "male" series of 21 cards, the authors found a rho of .75 between 28 male and 23 female college students. Since the multijudge reliability was .80, there appeared to be little sex difference with regard to the judgment of ambiguity. Dividing 15 cards, selected from the 21, into three groups of 5 cards each, according to their high, medium, or low ambiguity, they found that the medium-ambiguous series yielded themes of greater "personality revealingness" than either the low or highly ambiguous series. The difference in "revealingness" was not a function of the length of the story told. Dividing the 15 cards into the dichotomy used by Murray, 8 were found to belong to the "everyday" series, while 7 were assigned to the "fairy tale" series. No significant difference in "transcendence" was found between the two groups of cards. The interaction between ambiguity, "transcendence," and exposure time showed that the medium-ambiguity pictures yielded a greater amount of "transcendence" than the high-ambiguity ones under long exposure time (2 minutes) but not under short exposure time (5 seconds). The medium-ambiguity group of pictures did, however, show a higher "transcendence" than the least ambiguous pictures under both long and short exposure time.

Murstein (1958b), using Bijou and Kenny's (1951) rankings of ambiguity of a "male" series of cards with some modifications, found significant curvilinear correlations with the number of themes elicited from an all-male population composed of both college students and psychiatric patients used by Eron (1950). The moderately ambiguous cards produced the most themes, while the high- and low-ambiguous cards were less productive. While the curvilinear correlation involving the heterogeneous group and male-only group proved to be highly significant, the use of the female rankings of ambiguity did not yield a significant difference between eta and the Pearson r. The range of etas was from .43 to .95, the Pearson's r's ranging from .03 to .09.

In another study, Murstein (1958a) had 12 female *S*s rank a "female" series of 20 cards for "pleasantness" as well as for "psychological ambiguity." When compared to the "emotional tone" scores obtained by Eron (1953) from a similar female population, the following average correlations were obtained: ambiguity vs. emotional tone, .56; ambiguity vs. pleasantness, .40; pleasantness vs. emotional tone, .59. Thus, the results indicated that the more ambiguous the TAT card, the more "pleasant" the "emotional tone" of the story was found to be; the more "pleasant" the stimulus properties of the card, the more "pleasant" the "emotional tone" of the stories.

The results of the ambiguity experiments seem to show that the more ambiguous a picture is, the more pleasant are the themes elicited in the stories told to it. The more structured pictures are usually *negatively* structured (Murstein, 1958a) and, hence, more likely to elicit unpleasant themes. The medium-ambiguous pictures usually elicit more themes as well as more "personality revealing" material.

In appraising the work reviewed in this paper the following conclusions seem warranted:

Conclusions Derived From TAT Modifications

1. A lack of visual stimulation (darkness) seems to enhance the occurrence of pleasant association to TAT-type cards. The effect of supranormal illumination is more complex with a slight increase resulting in a manifestation of more pleasant responses than under normal lighting. As the intensity of illumination is further increased, however, the pleasantness of association diminishes.

2. Mere similarity between the *S* and the central figure in the stimulus card is insufficient to expedite the projection of meaningful responses. Projection is most readily elicited when the stimulus figure is culturally high in status and yet not so similar to the *S* as to arouse his suspicion as to the purpose of testing. It is not similarity per se that accounts for the censoring of responses, but the new awareness of the background characteristics (i.e., purpose of the testing) which causes the *S* to filter his responses. Other background factors such as the race of the examiner may also play a part in determining the nature of the response elicited.

3. The fact that Negro *S*s when tested by white examiners tend to be superficial in approach, hewing closely to mere description of the thematic cards, hardly means that they do not have tensions which could be percep-

tually expressed. Rather, the lack of involvement may have been a function of a low expectancy of satisfying these needs even to the extent of verbalizing them, because of the inhibiting presence of the examiner. With a Negro examiner, however, the stories told by Negroes would probably be more meaningful, though this particular comparison has not as yet been made.

The tendency of whites to give richer stories to the Thompson series rather than to the Murray TAT would seem to be a way of expressing either a need for verbalization of tolerance or a feeling of freedom in talking about a socially subservient group. Presumably, the Northern *S*s would be most prone to strive for tolerance in their stories, while Southern students (Light's group?) might be more apt to play upon the social and economic inferiority of Negroes.

4. Racial differences as determiners of projection have been overestimated, while socioeconomic factors have been underplayed. Middle-class Negroes may have more in common with middle-class whites with regard to perceived needs, than have middle-class Negroes with lower-class Negroes.

5. Animal pictures have been shown to be somewhat less effective than human pictures in eliciting meaningful projective responses from children.

Conclusions Derived From the TAT

1. Increasing the "psychological" ambiguity of a picture does not result in an increased response to the picture. Rather, the relationship is curvilinear with an increased response productivity being followed by a decrease as the gestalt of the picture begins to disappear.

2. The meaningfulness of stories for personality evaluation seems to follow a similar curvilinear relationship with ambiguity, with the medium-ambiguous pictures facilitating the expression of the most revealing stories.

3. The "pleasantness" of stories seems to increase as ambiguity increases. This occurrence may be a function of the fact that most highly structured cards on the TAT are negatively structured. Another factor may be that as the stimulus "pull" becomes weaker, the personality needs of the *S*, as well as the background factors such as the personality of the examiner, the reason the test is being given, and the locale of the test, become increasingly important. It seems reasonable to assume that most persons for cultural reasons prefer to see "good" rather than "bad." Also of importance for college students who form the hard core of *S*s used in these experiments is the need of impressing and favorably influencing the examiner who may well be perceived as a teacher-surrogate. Under these conditions, the exceedingly ambiguous picture may act as a cue to set the *S* on his guard to put his best perceptual foot forward.

4. The effect of varying the time of exposure is minimal provided the *S* may perceive the cards for at least a few seconds. Presenting the card for only a fraction of a second seriously affects the productivity of response. Allowing the picture to be exposed for two minutes has little if any advantage over showing it for five seconds.

5. The area of sex differences with regard to stimulus ambiguity is largely unknown. More information might be obtained if both sexes were given both the "male" and "female" series and comparisons made taking into consideration the sex of the figure depicted on the cards.

The knowledge that medium-ambiguous pictures produced the richest material personality-wise in the studies reviewed represents only a beginning in the understanding of the relationship between stimulus properties of the TAT and the responses educed. Medium-ambiguity may stem from the task of perceiving what is in the picture, or it may refer to the interpretation of what is clearly perceived. An inspection of the medium-ambiguity TAT cards leads to the observation that the objects in the pictures seem fairly clear; what is vague is the uncertainty as to the feelings the characters in the pictures seem to be experiencing. Another important feature of the medium-ambiguity category is the presence of humans on each of these cards. Thus, it would appear that the most powerful cards for eliciting idiosyncratic projective material are cards which contain humans who are fairly easily identified but whose expressions and feelings are capable of multiple interpretation. Thus, in Card 13 MF it is obvious to most persons that an apparently nude woman is lying beneath the covers of a bed, while a young man stands fully clothed with his arm covering his face. The ambiguity resides only in what the characters are doing and feeling. Maximum use ought to be made, therefore, of cards clearly depicting humans, whose behavior, however, is not readily discernible. Such cards would appear to have more appeal for *S*s than fantastic "fairy tale" type pictures, at least with regard to normal populations.

In reviewing the earlier studies, the importance of stimulus control and background factors has been shown to play crucial roles in the elicitation of responses to TAT-type techniques. It would appear that the emphasis upon the TAT's ability to tap the "private world" of the individual has hindered clinicians from fully utilizing other equally meaningful determinants of perception. In fact, one of the key repeating factors seeming to run through the studies covered has been the fact that this "private world" has been revealed only under rather specific conditions and with specific cards. One must conclude, therefore, that much of the thematic material educed is obtained with the permission of the *S*, and is subject to his control. This statement is in accordance with Lindzey and Tejessy's finding (1956) that TAT hostility indices correlated more highly with the *S*s self-concept than with observer or clinical ratings. Moreover, Allport (1953) has said, ". . . normal subjects . . . tell you by the direct method precisely what they tell you by the projective method. They are all of a piece. You may therefore take their motivation statements at their face value, for even if you probe, you will not find anything substantially different" (p. 110).

If the *S*s' control of the testing situation is so pervasive, how can we, as clinicians, design tests so as to circumvent this control and obtain meaningful glimpses into that "private world" which is not offered for public consumption? The earlier work seems to indicate that medium-ambiguous pictures should be used where the figures are humans, culturally acceptable as objects of identification and yet not so similar to the *S* as to cause him to become overly self-conscious and heavily censor his responses.

Finally, the task of quantifying the stimulus, background, and organismic variables is a formidable one. The stimulus value of personality variables such as aggression and sex, has been scaled for TAT-type pictures, using the Guttman scaling technique (Auld, Eron, and Laffal, 1955). Examiner attitudes and behavior could also be measured through an amalgam of scaling procedures and observation and personality tests, as could the behavior of the *S* apart from the testing situation. The quantification of these variables would lend meaning to the responses obtained via the projective techniques. Currently, for example, one is often at a loss to determine whether responses on the TAT reflecting hostility are part of the *S*'s perceptual world or merely reflect a respect for the objective properties of the cards. By means of the Stimulus × Background × Organismic approach, however, such questions may be answered as well as ones regarding the relationship between responses and overt behavior. Such an approach may well lead to a more homogeneous thematic theme test whose stimulus values are known for various personality dimensions with regard to different populations, thus mitigating the manifold problems of interpretation.

Received January 9, 1958.

REFERENCES

ALLPORT, G. W. The trend in motivational theory. *Amer. J. Orthopsychiat.*, 1953, **23**, 107–119.

ARMSTRONG, MARY A. S. Children's responses to animal and human figures in thematic pictures. *J. consult. Psychol.*, 1954, **18**, 67–70.

AULD, F., ERON, L. D., & LAFFAL, J. Application of Guttman's scaling method to the T. A. T. *Educ. psychol. Measmt*, 1955, **15**, 422–435.

BEIER, E. G., IZARD, C. E., SMOCK, C. D., & TOUGAS, R. R. Response to the human face as a standard stimulus: A re-examination. *J. consult. Psychol.*, 1957, **21**, 165–170.

BELLAK, L., & BELLAK, SONYA S. *Manual of instruction for the Children's Apperception Test.* New York: C.P.S. Co., 1949.

BIERSDORF, KATHRYN R., & MARCUSE, F. L. Response of children to human and to animal pictures. *J. proj. Tech.*, 1953, **17**, 455–459.

BIJOU, S. W., & KENNY, D. T. The ambiguity of TAT cards. *J. consult. Psychol.*, 1951, **15**, 203–209.

Bills, R. E. Animal pictures for obtaining children's projections. *J. clin. Psychol.*, 1950, **6**, 291–293.

Bills, R. E., Leiman, C. J., & Thomas, R. W. A study of the validity of the TAT and a set of animal pictures. *J. clin. Psychol.*, 1950, **6**, 293–295.

Bradley, J. E., & Lysaker, R. L. Ambiguity as a variable in the use of a projective technique. Pillsbury Mills, Inc., Minneapolis, Minn., 1957. (Mimeo.)

Brozek, K., Guetzkow, H., & Baldwin, M. V. A quantitative study of perception and association in experimental semistarvation. *J. Pers.*, 1951, **19**, 245–264.

Cook, R. A. Identification and ego defensiveness in thematic apperception. *J. proj. Tech.*, 1953, **17**, 312–319.

Edgar, Clara L., & Shneidman, E. S. Some relationships among thematic projective tests of various degrees of structuredness and behavior in a group situation. *J. proj. Tech.*, 1958, **22**, 3–12.

Eron, L. D. A normative study of the Thematic Apperception Test. *Psychol. Monogr.*, 1950, **64**, No. 9 (Whole No. 315).

Eron, L. D. Responses of women to the Thematic Apperception Test. *J. consult. Psychol.*, 1953, **17**, 269–282.

Frank, L. K. *Projective methods.* Springfield, Ill.: Charles C Thomas, 1948.

Furuya, K. Responses of school children to human and animal pictures. *J. proj. Tech.*, 1957, **21**, 248–252.

Goodman, Mary E. *Race awareness in young children.* Cambridge, Mass.: Addison-Wesley, 1952.

Helson, H. An experimental approach to personality. *Psychiat. Res. Rep.*, 1955, No. 2, 89–99.

Kenny, D. T. Transcendence indices, extent of personality factors in fantasy responses, and the ambiguity of TAT cards. *J. consult. Psychol.*, 1954, **18**, 345–348.

Kenny, D. T., & Bijou, S. W. Ambiguity of pictures and extent of personality factors in fantasy responses. *J. consult. Psychol.*, 1953, **17**, 283–288.

Korchin, S. J., Mitchell, H. E., & Meltzoff, J. A critical evaluation of the Thompson Thematic Apperception Test. *J. proj. Tech.*, 1950, **14**, 445–452.

Lasaga y Travieso, J. I., & Martinez-Arango, C. Some suggestions concerning the administration and interpretation of the TAT. *J. Psychol.*, 1946, **22**, 117–163.

Levine, R., Chein, I., & Murphy, G. The relation of the intensity of a need to the amount of perceptual distortion: A preliminary report. *J. Psychol.*, 1942, **13**, 283–293.

Light, B. H. Comparative study of a series of TAT and CAT cards. *J. clin. Psychol.*, 1954, **10**, 179–181.

Light, B. H. A further test of the Thompson TAT rationale. *J. abnorm. soc. Psychol.*, 1955, **51**, 148–150.

Lindzey, G., & Tejessy, Charlotte. Thematic Apperception Test: Indices of aggression in relation to measures of overt and covert behavior. *Amer. J. Orthopsychiat.*, 1956, **26**, 567–576.

Mainord, Florence R., & Marcuse, F. L. Responses of disturbed children to human and to animal pictures. *J. proj. Tech.*, 1954, **18**, 475–477.

Murray, H. A. *Thematic Apperception Test Manual.* Cambridge: Harvard Univer. Press, 1943.

Murstein, B. I. Nonprojective determinants of perception on the TAT. *J. consult. Psychol.*, 1958a, **22**, 195–198.

Murstein, B. I. The relationship of stimulus ambiguity on the TAT to productivity of themes. *J. consult. Psychol.*, 1958b, **22**, 348.

Riess, B. F., Schwartz, E. K., & Cottingham, Alice. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, **45**, 700–709.

Rotter, J. *Social learning and clinical psychology.* New York: Prentice-Hall, 1954.

Schwartz, E. K., Riess, B. F., & Cottingham, Alice. Further critical evaluation of the Negro version of the TAT. *J. proj. Tech.*, 1951, **15**, 394–400.

Thompson, C. E. The Thompson modification of the Thematic Apperception Test. *Rorschach Res. Exch.*, 1949, **13**, 469–478.

Tomkins, S. S. *The Thematic Apperception Test.* New York: Grune & Stratton, 1947.

Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950a, **29**, 379–390.

Weisskopf, Edith A. An experimental study of the effect of brightness and ambiguity on projection in the Thematic Apperception Test. *J. Psychol.*, 1950b, **29**, 407–416.

Weisskopf, Edith A., & Dunlevy, G. P. Bodily similarity between subject and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.*, 1952, **47**, 441–445.

Weisskopf-Joelson, Edith A., & Lynn, D. B. The effect of variations in ambiguity on projection in the Children's Apperception Test. *J. consult. Psychol.*, 1953, **17**, 67–70.

Weisskopf-Joelson, Edith A., & Money, L. Facial similarity between subject and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.*, 1953, **48**, 341–344.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

MEASUREMENT OF IMPROVEMENT IN "MENTAL ILLNESS"[1]

ROBERT B. ELLSWORTH[2] AND WILLIAM H. CLAYTON
Veterans Administration Hospital, Salt Lake City, Utah

The concept of mental illness as a disease process has resulted in the focusing of psychiatric attention on the symptoms which a "mentally ill" patient manifests. The extent to which a mentally ill person improves is traditionally based on the extent to which the original symptom picture becomes modified. As with most disease entities, the patient is judged as having recovered when the symptoms of the disease abate.

Descriptive psychiatry, as practiced in most mental hospitals today, is based on the diagnostic classifications outlined by Kraepelin. His observation that the praecox progressed to a demented end-state led him to believe that this process was based on some degenerative "disease" process. Later evidence discredits the idea of the necessary demented end product, and, to a lesser extent, that this disease occurs only in young adults. The original evidence, then, for regarding dementia praecox as a disease process is no longer consistent with observations following those of Kraepelin. The hypothesis may still be true, but the evidence upon which the hypothesis was based is no longer accepted.

The agreement of the American Psychiatric Association on a classification system in 1934 was a major step forward. Professional personnel had at last come to agree on a common system which facilitated communication and research. It is unfortunate, however, that many psychiatrists still regard the classifications as synonymous to disease entities, although the majority of the psychiatric group regards the classifications as reaction patterns rather than disease entities. This is the current status of the Veterans Administration's nomenclature, adopted in 1952.

The importance of the behavior patterns in the schizophrenic reactions was pointed out by Adolf Meyer. He and his students, notably Norman Cameron, emphasized the influence of social factors in the development, onset, and form of these reactions. Indeed, Cameron (1947) prefers to include the schizophrenics under a broad classification of Behavior Disorders. The implications which arise from the use of "disease," "symptoms," "mental illness" are thus avoided by Cameron.

The assessment of behavioral adjustment offers another approach in the evaluation of psychiatric patients. Many psychiatric patients who exhibit a minimum of psychiatric symptomatology are unable to leave a mental hospital and lead productive lives. Other patients, despite extensive psychopathology, leave mental hospitals and become productive citizens. The extent of psychopathology, therefore, is not necessarily related to the adequacy of social functioning, although there is generally a high incidence of behavioral maladjustment in severely "ill" psychiatric patients.

This paper, then, presents some data on the relationship between behavioral adjustment and psychopathology as related to each other and to the level of posthospital adjustment. Perhaps the most stringent criterion of any measure of adjustment in psychiatric patients is the relationship between such a measure and posthospital adjustment. Such an analysis is also attempted in this paper, relating level of behavioral adjustment and psychopathology at time of discharge to the level of social adjustment three months following discharge.

---

[1] In cooperation with Psychiatric and Nursing Services, particularly Anna Kontas, who supervised the collection of all admission behavioral data. This study was supported by the Psychiatric Evaluation Project, Richard L. Jenkins, Director.

[2] Now at VA Hospital, Ft. Meade, South Dakota.

Procedure

Between August 1956 and May 1957, every functional psychotic patient under 60 years of age, without major physical illness, having been hospitalized no longer than 90 days during the preceding 6 months, admitted to the Fort Douglas VA Hospital, became a subject in this study. The patient was interviewed upon admission and rated (RBE) on a psychopathology scale. A psychiatric nurse and aide rated him independently on a behavioral adjustment scale following three days' observation. All personnel using the behavioral adjustment scale were trained raters.

For purposes of this study, part of this admission group ($N = 78$) was broken into two additional subgroups for further study: 1. Those who were discharged from the hospital within 6 months ($N = 27$). The patients in this group were interviewed and again rated on the psychopathology scale at the time of discharge, and were also rated independently by a psychiatric nurse and aide on the behavioral adjustment scale at this time. Twenty-five of the 27 patients in this group were personally interviewed 3 months following discharge (WHC) and rated on a Posthospital Adjustment Scale (described below). These 25 included all patients who had been contacted and interviewed at the time of this analysis. 2. Those who remained in the hospital 6 months after their admission ($N = 11$). All rating scales were also completed on this group. The remaining 40 patients who were in neither of the subgroups were: (*a*) those whose exit ratings could not be completed (no trained aide or nurse rater) and elopements; (*b*) those who were neither discharged nor had resided in the hospital for a period of 6 months at the time of this analysis.

Psychopathology Scale

A modified Lorr Multi-dimensional Rating Scale (1953) was used in this study. The Lorr scale was modified to meet the needs of the Psychiatric Evaluation study and subjected to extensive reliability analysis. Several items were dropped if they failed to reach an acceptable degree of interrater reliability (13 raters independently rating 52 patients). The items that remained included manifest and reported depression, manifest and reported anxiety, apathy, reported hallucinations, delusions or morbid suspicion, thought disorder, rapport, disorientation, bizarre posturing, memory deficit, lack of motivation, excessive hostility, withdrawal, evasion, and cooperativeness. These items represent, then, a sampling of Lorr's psychopathology factors.

Seven psychiatrists were asked to judge: "at what level (rating) each symptom would have to be manifest in an interview before you would be confident of the presence of underlying psychopathology." The level at which at least 6 out of 7 agreed, was accepted as the psychopathology level. All 7 agreed completely on 74% of the items. Each level beyond this judgment was scored 1 point toward a Total Psychopathology score. Seventeen items (described above) were judged as indicating psychopathology. On these 17 items, a patient could score 31 pathology points, since most items contained 2 levels (ratings) beyond the judged point of psychopathology. A Total Psychopathology score was thus obtained; the higher the score, the more extensive the psychopathology. Along with the author one additional rater independently rated 17 patients with a resultant rho of .84 between the Total Psychopathology scores for the 2 raters. The reliability of the Total Psychopathology score was thus found to be within acceptable limits.

An intercorrelational analysis of all psychopathology items (phi correlation, N of 78 ratings) revealed that each item was "relatively" independent of every other item in the scale. Although several items were significantly interrelated, the highest phi was .51, which accounted for only 25% of the variance between the 2 items. The relative independence of each of the items permitted the scoring of a Total Psychopathology score, with each item contributing to the total score.

Behavioral Adjustment Scale

The MACC Behavioral Adjustment Scale (1957) was used in this study. The development, reliability, and validity outlined in the manual are not reported here. The scale measures 4 relatively independent behavioral clusters: Motility, Affect, Cooperation, and Communication. The latter 3 areas constitute the Total Adjustment score. (See Ellsworth, 1957.)

The ratings of 2 independent raters (nurse and aide) were averaged, thus decreasing the interrater error. The interrater reliability of the 2 admission raters (Total Adjustment scores, 78 paired ratings) was .89.

Posthospital Adjustment Scale

As part of the Psychiatric Evaluation project,[3] 25 psychiatric patients were followed into the community and rated in 6 areas of social adjustment. These included occupational (or school) adjustment, management of funds, family adjustment, "other" interpersonal relationships, recreational adjustment, and community adjustment. One of the authors (Clayton) developed a 5-point rating scale for each of these posthospital adjustment areas. For example, a patient who is employed full time receives a rating of 1 in the area of occupational adjustment, while a patient who refuses to seek work and does not help at home is rated 5. A low total score (for the 6 areas) thus indicates "good" posthospital adjustment, while a high score indicates "poor" posthospital adjustment.

A social worker (Clayton) visited the patient and his family 3 months following discharge and rated each patient in the 6 posthospital adjustment areas. He was not aware of either the patient's behavioral adjustment or pathology ratings at the time of the interview and rating. Following the interview, he wrote a narrative report describing each patient's social situation and conditions under which he was living, including the behavioral patterns and relationships in each of the 6 areas of adjustment. Two additional social workers independently rated all 25 patients from this narrative report, with a resulting average intercorrelation of .82 between all raters on the total scores. Thus, the Posthospital Adjustment Scale appears to result in satisfactorily reliable scores which permit the scoring of each patient's over-all posthospital adjustment.

Results

a. Table 1 presents the intercorrelations between Total Psychopathology scores and Behavioral Adjustment areas upon admission. The results indicate that the higher the behavioral adjustment, the lower the extent of psychopathology. In other words, the more pleasant, cooperative, and communicative the patient was, the *less* apt he was to show extensive psychopathology. The relatively high intercorrelations between two different kinds of scales, rated independently, are somewhat surprising. This would seem to suggest that these two assessment approaches of the adjustment of the newly admitted patient with a psychotic reaction, whether it be from a behavioral or psychopathological standpoint, have much in common.

b. The amount of improvement (score change) on both behavioral adjustment and psychopathology was analyzed. The score changes of Subgroup 1 (those who were discharged within 6 months of admission) were compared on both scales. This resulted in a product-moment correlation of .74 (significant beyond .01 level of confidence). Thus, the discharged patient who improves in his behavioral adjustment tends to show a similar amount of improvement in his psychopathology symptoms. Both methods of assessing improvement have much in common, as would be expected from the results of Table 1.

c. The results of Subgroup 2 (those who were not seen as ready for discharge at the end of 6 months of hospitalization) are summarized in Table 2. The mean scale changes of this "in-hospital group" were compared with the mean scale changes of those actually discharged before reaching the 6-month point of hospitalization (Subgroup 1).

It appears that, with regard to using discharge as a criterion of "improvement," the psychopathology scores do not distinguish between those improved sufficiently to warrant discharge within 6 months and those who were seen as needing hospitalization beyond 6 months. The behavioral adjustment scale scores, however, reflected a significant change in the "improved" (discharged) group,

---

[3] This VA project, to evaluate the effectiveness of treatment of several VA hospitals, is directed by Richard L. Jenkins and Lee Gurel.

Table 1

Intercorrelations Between MACC Behavioral Adjustment Scores and Total Psychopathology Score Upon Admission

($N = 78$)

	Motility	Affect	Cooperation	Communication	Total Adjustment
Total psychopathology	+.49*	−.51*	−.70*	−.65*	−.69*

* Significant beyond .01 level of confidence. Motility scores are not included in the Total Adjustment scores.

but not in the "unimproved" group. In so far as this particular VA hospital operates, behavioral adjustment appears to be more highly related to improvement (as measured by discharge) than psychopathology.

d. A fourth analysis involved the relationship between the length of hospital stay with behavioral adjustment level and total psychopathology score at the time of admission. In other words, which is more predictive of the length of hospital stay (which is an indirect measure of the severity of the psychiatric disorder), the extent of psychopathology or the level of behavioral adjustment at the time of admission? With 35 patients who had been discharged from the hospital at the time of this analysis (27 discharged before, 8 discharged after 6 months of hospital treatment), a product-moment correlation of .21 (not significant) between extent of psychopathology and length of hospital stay, and a correlation of − .41 (significant at .01 level) between behavioral adjustment level and length of hospital stay resulted. It appears that the patient who has the highest behavioral adjustment upon admission tends to remain in the hospital the shortest time. The extent of pathology symptoms on admission does not appear to be significantly related to length of hospital stay.

e. The last analysis involved the relationship between extent of psychopathology and level of behavioral adjustment at time of discharge, to posthospital adjustment three months following discharge. Thus, it was possible to relate these two measures of the adjustment of psychiatric patients to the adequacy of posthospital community functioning. The effectiveness of hospital treatment is perhaps best measured against the adequacy with which an ex-patient handles his life following discharge.

In this analysis, the product-moment correlation between extent of psychopathology upon discharge and level of posthospital adjustment was .22 (not significant, $N = 25$). The level of behavioral adjustment at time of discharge, in relation to level of posthospital adjustment, was − .47 (significant beyond the .02 level of confidence). Thus, the behavioral adjustment, as rated independently by a psychiatric aide and nurse at time of discharge, was significantly related to level of community adjustment 3 months following discharge; since the patient who adjusted well in the community (low score) tended to

Table 2

Mean Behavioral Adjustment and Psychopathology Changes of Patients Not Seen as Ready for Discharge Compared with Those Discharged Within Six Months of Hospitalization

Groups	*N*	Behavioral Mean		*t*	Pathology Mean		*t*
		1[a]	2[b]		1[a]	2[b]	
Discharged within 6 months	27	40.3	46.7	3.85*	6.2	1.4	5.92*
Not discharged within 6 months	11	39.3	43.0	1.20	6.3	1.9	3.3*

* Significant beyond .01 level of confidence.

[a] Refers to the mean of behavioral and pathology scores at time of hospital admission.

[b] Refers to the mean of behavioral and pathology scores at time of discharge from the hospital.

have the *highest* behavioral rating at the time of discharge.

Discussion

Although behavioral adjustment and psychopathology might, at first glance, appear to be two entirely different approaches in evaluating psychiatric patients, they are both based on behavioral observations. Symptomatology might be described as abstractions from behavior. For instance, two "delusional" patients may differ widely in their behavior in relation to a delusion. Behavioral adjustment is the more concrete description of behavior, while psychopathology is the more abstract description. This explains in part their similarity as demonstrated in Results *a* and *b*.

A major difference between behavioral adjustment and psychopathology, however, which may partly explain the superiority of behavioral adjustment in relation to Results *c*, *d*, and *e*, is that psychopathology assessment has no "top." Psychopathology scales range from the presence of symptoms to their absence. Psychopathology does not go beyond the absence of symptoms, i.e., two patients who show no delusions, thought disorganization, etc., may be very different people. One may be friendly and out-going while the other may be reserved and withdrawn. Behavioral adjustment, on the other hand, extends the assessment into measuring both the degree of positive as well as negative behaviors.

The analyses of the relationships of psychopathology and behavioral adjustment to such "external" criteria as the length of hospitalization (indirectly a measure of the severity of psychiatric condition), readiness for discharge (Result *d*), and level of social adjustment three months after discharge (Result *e*) revealed that an assessment of traditional psychopathology symptoms was not significantly related to any of these external criteria. Yet, in most psychiatric hospitals, the "improved" patient is the patient whose symptoms have cleared even though such improvement may not be related to how well the patient handles his life after he leaves the hospital. The present study is limited by the inclusion of only functional psychotic patients and by the choice of only those psychopathology symptoms which were found to be acceptably reliable in an earlier study. Nevertheless, the present study would seem to raise some serious doubts as to the usefulness of psychopathology as a meaningful criterion of improvement in relation to other external criteria of adjustment. The assessment of behavioral adjustment as a measure of improvement would seem to offer a more promising approach. As was seen in Result *c*, patients often show a significant decrease in psychopathology without developing a sufficiently high level of behavioral adjustment to warrant discharge. The relative absence of psychopathology, then, does not necessarily mean that the patient is an adequate social individual. On the other hand, many individuals with rather extensive psychopathology are able to handle themselves quite adequately in everyday social situations.

It is of more than historical interest to note that the era of "moral" treatment in the early 19th century resulted in over 75% of first admission psychiatric patients released as cured or improved, and over 50% of these patients never relapsing up to 36 to 60 years following release (Bockoven, 1956). These results, which would be looked upon with pride by most psychiatric hospitals today, occurred before the advent of well defined systems of psychopathology, and certainly before the psychotic process was looked upon as a disease entity. Bockoven (1956) concludes that these earlier recovery rates were not a result of a given treatment but a result of the nonobstruction of the natural course of mental illness. Recovery occurs when patients are treated as human beings, and, "The moral therapist acted toward his patients as if they were mentally well . . . in impressing on patients the idea that a change to more acceptable behavior was expected" (Bockoven, 1956, pp. 302–303). Clearly an emphasis was placed on behavior; and consequently, as Bockoven suggests, this early approach to treatment of mental illness must seem naive and superficial to modern dynamic psychiatry. The efforts of many 20th century psychiatrists, however, to establish mental illness as a disease entity and to bring this field closer to medicine probably resulted, until recently, in an increased psychological distance from the patient with consequent neglect of him as a human being.

Conclusions

1. Measures of behavioral adjustment and psychopathology are significantly related, warranting the use of both as measures of improvement in mental illness.

2. Behavioral adjustment appears to be more sensitive to "patients ready for discharge," in predicting length of hospital stay and in predicting the level of posthospital adjustment, than psychopathology. The lack of any significant relationship between psychopathology assessment and these "external" criteria would seem to raise some serious doubts as to the usefulness of psychopathology as a socially meaningful criterion of improvement in functionally psychotic patients.

Received January 9, 1958.

REFERENCES

Bockoven, J. S. Moral treatment in American psychiatry. *J. nerv. ment. Dis.*, 1956, **124**, 167–194, 292–321.

Cameron, N. *The psychology of behavior disorders.* Boston: Houghton Mifflin, 1947.

Lorr, M. Multi-dimentional psychiatric rating scale. Washington, D. C.: *VA Tech. Bull.* 10-507, 1953.

Ellsworth, R. B. The MACC Behavioral Adjustment Scale. Los Angeles: Western Psychological Services, 1957.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

A VALIDATION STUDY OF THE SELF-ACTIVITY INVENTORY AS A PREDICTOR OF ADJUSTMENT

NELDA S. NEBERGALL, HENRY ANGELINO, AND HARL H. YOUNG
University of Oklahoma

The Self-Activity Inventory (SAI) developed by Philip Worchel (1957) for the U. S. Air Force was designed as a psychiatric screening device as well as to aid in understanding self-concept variables. The SAI was originally devised to test the susceptibility to failure of pilot candidates under conditions of military stress. Subsequently, some SAI variables have been shown to be related to decrements in performance under stress as applied in the laboratory (Cawthon, Oldroyd, & Young, 1957; Miller & Worchel, 1957).

Moreover, the SAI has been utilized in several investigations with both males and females (Cawthon et al., 1957; Miller & Worchel, 1957; Wilkinson, 1956; Young, 1956), *but* previous validation studies were conducted with males only. Because some of the above studies, and others by Witkin (1954), revealed interactions between sex and other variables as well as actual sex differences, the necessity for additional validation which would take into account the sex variable was indicated.

The present study was undertaken to examine relationships between self-concept variables as measured by the SAI and criterion variables consisting of group-rankings and self-rankings of personal adjustment, and to investigate possible sex differences among these variables.

Method

Population. The population consisted of 50 female and 44 male students. Each subject had resided in his or her fraternity or sorority house for at least five months. It was assumed that each was well acquainted with all other students in the group due to the intimate contact provided by group activities and general proximity. The participating college students were no longer living in their family groups but were required to accept new forms of discipline and authority relationships, to form new positive identifications in a context different from that of the family, and to forego narcissistic need for self-gratification in favor of group standards and goals. These are criteria of personal adjustment mentioned by Worchel (1957) in the rationale of the SAI.

Procedure. Identical procedures were followed in collecting data for males and females. A like-sex experimenter worked with each group. All subjects were given the SAI before being asked to perform the ranking procedure.

Description of the SAI. The SAI is a self-rating scale consisting of 54 statements describing responses of hostility, achievement, sex, and dependency in particular situations. The scale was reproduced in two identical forms for males and females, except that personal pronouns appeared in the appropriate gender on each form. The subject makes three ratings on a five-point scale for each item, according to the following scheme which appears as three columns on the test blank: "I am a person who . . . ," "I would like to be a person who . . . ," "The average person is one who. . . ." Summing the ratings in each of the three columns yields a Self score (S), an Ideal Self score (I), and a score for the "Average Person" (A). The items are so worded that high scores on each of these variables indicate a trend toward depreciation on the variable in question; that is, the higher a person's Self score the lower is his self-esteem,

Bureau of [illegible] & Psyl. Research
S. C. E. R. T.)
Date 6.3.81

and similarly, the higher his Ideal-Self score the greater is the tendency to depreciate his ideal self, and so on. In addition to these three scores, the absolute discrepancies between ratings can be summed across the 54 items yielding three discrepancy scores: the Self-Ideal (S-I), Ideal-Average (I-A), and Self-Average (S-A) discrepancy scores. Absolute discrepancy scores are used because algebraic discrepancy scores derived from this instrument have mathematical properties which severely limit their usefulness (Worchel, 1956).

Description of the ranking procedure. Following completion of the SAI, each subject was given a list of members in his organization and asked to rank each individual in the group, including himself, on "personal adjustment." The technique was similar to that used by Holtzman (1950a; 1950b), and Calvin and Holtzman (1953).

Personal adjustment was defined for the subjects as follows:

> An ideal personal adjustment by which you might judge would be the person who is free from emotional problems, who handles personal problems well and who rarely shows signs of excessive worry or anxiety. The person functions with a minimal amount of conflict and tension in relation to the rest of the group and has harmonious interpersonal relationships. This person seems to be basically happy and is liked by others.

Table 1

Correlations Among Self Rank, Group Rank, and SAI Scores

Rating	SAI	Boys ($N = 44$)	Girls ($N = 50$)
		r	r
GR	X Self	−.15	−.01
GR	X Ideal	−.04	−.07
GR	X Average	−.00	−.22
GR	X S–I	−.11	−.13
GR	X S–A	.05	.04
GR	X I–A	−.13	−.07
SR	X Self	.47**	.27
SR	X Ideal	.16	.06
SR	X Average	.10	.09
SR	X S–I	.49**	.33*
SR	X S–A	−.06	−.03
SR	X I–A	.18	−.00
SR–GR	X Self	−.39**	−.15
SR–GR	X S–I	−.51**	−.39**
SR	X GR	.22	.21

* Significant at the .05 level of confidence.
** Significant at the .01 level of confidence.

This description of personal adjustment specifically mentions factors related to both personal adjustment and to interpersonal or group adjustment.

The group-ranking score for each individual on personal adjustment was the mean of individual ranks assigned to a person by his associates. The group-ranking for a member thus represented the consensus of his associates. The self-ranking represented an individual's opinion of himself on the adjustment variable. Thus, two independent estimates of personal adjustment were obtained. A high numerical rank in both instances is depreciating, while a low numerical rank is enhancing. In addition, the absolute difference between group rank (GR) and self rank (SR) provided self-group rank (SR-GR) discrepancy scores. The use of the absolute difference score (SR-GR) was justified because in this study it is assumed that personal adjustment is in part a function of divergences between self and group evaluations of the individual. Thus the magnitude of discrepancy, rather than direction, is the important consideration.

Results and Discussion

Correlations of SAI variables with self and group ranks are presented in Table 1.

Examination of Table 1 indicates that group rank did not correlate significantly with any SAI variables. Self rank (SR) and the discrepancy between self and group rank (SR-GR) were significantly correlated only with the S and S-I scores from the SAI. Such significant relationships were present more frequently for men. However, Fisher's z transformation applied to pairs of correlations for males and females revealed that none of the scores was significantly different.

Thus attempts to validate the SAI as a predictor of adjustment were uniformly unsuccessful under the conditions of this study. The most meaningful and reliable criterion of adjustment (GR) showed no significant relationship with the SAI scores, nor any pronounced trends in this direction. This finding points to the limitations of the SAI when used for this purpose.

Since SR and S are two independent estimates of presumably similar aspects of a person's self-concept, a positive correlation between these two variables was expected. This anticipated relationship was obtained with a significance level of .01 for males. The .27 value of *r* for females very closely approximates the .28 required for a significant correlation at the .05 level of confidence with 48 *df*. It was not surprising the S-I was also significantly correlated with SR, since S and S-I were not derived independently. This relationship holds at the .01 level for males and .05 level for females.

SR and GR correlated .22 and .21 for males and females, respectively. There appears to be little relationship between how the individual views himself and the group consensus of him on personal adjustment. It is also interesting to note that the groups of both sexes ranked the large majority of individuals lower (less well-adjusted) than these persons ranked themselves. In other words, individuals and groups in this college population did not agree on the degree of adjustment of the individuals, and a prime source of this disagreement arose from the fact that most subjects saw themselves as better adjusted than did the group.

Probably the most striking findings of this study, however, center around the relationships between the SR-GR scores and the S and S-I variables. The value −.39 was obtained between the S and SR-GR scores for males. This *negative* relationship suggests that if the individual, on the SAI, tends to express freely what one may term socially disapproved feelings and behavior, his judgment of his personal adjustment tends to be similar to the judgment of the group. Conversely, if he denies such feelings and behavior, his judgment tends to disagree with the group consensus. It may be that this denial is one factor that the individuals in the group use in making global evaluations of adjustment. However, this inference is tempered by the fact that the corresponding correlation for females was not significantly different from zero.

Significantly *negative* correlations between SR-GR and S-I of −.51 and −.39 were obtained for males and females, respectively. Thus a small self-ideal discrepancy is associated with a large divergence between self and group judgments of adjustment and vice versa. At first glance one might conclude that this result casts some doubt on the contentions of many "self" theorists (Lecky, 1951; Rogers, 1951; Snygg & Combs, 1949) that self-ideal congruence is positively associated with adjustment. It should be remembered that whereas self-ideal discrepancy is a measure of discrepancy between two aspects of the individual's self-perception, SR-GR is a measure of discrepancy between the individual's self-perception and the perception of the individual by others. A relatively large or small self-ideal discrepancy was found to be associated with disagreement, *regardless of direction,* between the group and individual estimates of adjustment. Thus a person with a small self-ideal discrepancy tends to rank himself either considerably above *or* below the group rank for himself. On the other hand, the individual who is more openly dissatisfied with himself (large self-ideal discrepancy) tends to give himself a ranking on adjustment appreciably more in accord with that of the group.

Summary and Conclusions

In attempting to assess the validity of the Self-Activity Inventory with a mixed sex population, all resident members of a sorority and a fraternity were administered the SAI and asked to rank themselves, along with all members of their respective groups, on "personal adjustment." Scores were obtained for each subject on six variables from the SAI and were correlated separately for each sex with self rank and with pooled group rank. Self scores and Self-ideal discrepancy scores were correlated with a discrepancy score consisting of the absolute difference between self and pooled group ranking.

The results indicated that SAI scores were not related to the major criterion of adjustment used in this study, namely, pooled group rank. There were no significant differences in the magnitudes of the correlations for the sexes between SAI scores and either pooled group rankings or the self-rank group-rank discrepancy scores. Under the conditions of this study the validity of the SAI proved to be markedly limited.

The most provocative finding showed that a relatively large or small self-ideal discrepancy is significantly associated with disagreement, *regardless of direction,* between the group and individual estimates of adjustment. The results also suggested that the individual who tends to give free expression to self-depreciative attitudes on the SAI is more likely to agree with the group judgment of his personal adjustment than one who tends to deny such attitudes. In general, there was little relationship between self and group ranks of the individual's adjustment. However, inspection revealed that most people ranked themselves above the group evaluation, i.e., more well-adjusted.

Received January 9, 1958.

REFERENCES

CALVIN, A. D., & HOLTZMAN, W. H. Adjustment and discrepancy between self concept and inferred self. *J. consult. Psychol.,* 1953, **17**, 39–44.

CAWTHON, NELDA, OLDROYD, C. R., & YOUNG, H. H. Task performance under stress as a function of self-ideal discrepancy. Paper read at Southwestern Psychological Association, April, 1957.

HOLTZMAN, W. H. Validation studies of the Rorschach test: Impulsiveness in the normal superior adult. *J. clin. Psychol.,* 1950a, **6**, 348–351.

HOLTZMAN, W. H. Validation studies of the Rorschach test: Shyness and gregariousness in the normal superior adult. *J. clin. Psychol.,* 1950b, **6**, 343–347.

LECKY, P. *Self-consistency: A theory of personality.* New York: Island Press, 1951.

MILLER, KENT S., & WORCHEL, PHILIP. The effects of need-achievement and self-ideal discrepancy on performance under stress. *J. Pers.,* 1957, **25**, 176–190.

ROGERS, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.

SNYGG, D., & COMBS, A. W. *Individual behavior: A new frame of reference for psychology.* New York: Harper, 1949.

WILKINSON, A. E. Self consistency as a function of self-concept of parents and self-concept of children. Unpublished doctoral dissertation, Univer. of Texas, 1956.

WITKIN, H. A. *Personality through perception.* New York: Harper, 1954.

WORCHEL, P. Adaptability screening of flying personnel. Development of a self-concept inventory for predicting maladjustment. *USAF Air Univer. Sch. Aviat. Med.,* Randolph Air Force Base, Texas, Rep. No. 56-62, 1957.

YOUNG, H. H. Sex differences in the self-concept and related variables. Paper read at Oklahoma State Psychological Association, November, 1956.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

THE EFFECTIVENESS OF CLINICIANS' JUDGMENTS:

THE DIAGNOSIS OF ORGANIC BRAIN DAMAGE FROM THE BENDER-GESTALT TEST[1]

LEWIS R. GOLDBERG[2]
University of Michigan

One of the most penetrating criticisms of previous attempts at assessing clinical judgments has been that the experimental predictions asked of the clinician differed in class or in kind from those made in his day-to-day practice. More specifically, the clinician often felt that the prediction of success at flight school (Holtzman & Sells, 1954), competency in clinical psychology (Kelly & Fiske, 1951), or even outcome of psychotherapy (Barron, 1953), called for a range of inferences beyond those usually demanded of him. Some clinicians deny making predictions and, instead, characterize their usual work in terms of "personality description," "diagnosis," "dynamic formulations," or "understanding the patient." Seen in this light, a good many of the published studies of the clinician's effectiveness can be dismissed as irrelevant by practicing clinicians who see themselves primarily as diagnosticians or therapists.

To eliminate these objections and thereby place clinicians in their best light, clinical assessment must: (*a*) involve problems typically encountered by the practicing clinician (for example, diagnostic judgments of neuropsychiatric referrals), (*b*) allow the clinician to use his favorite techniques (be they tests, interviews, case history material, reports from other services, etc.) in his favorite manner of utilizing them, and then (*c*) independently validate his conclusions against evidence acceptable to science as a whole; moreover, this entire procedure should be compared with similar judgments made by nonpsychologists (for example, clerical help).

Once one has established the kinds of judgments which clinicians tend to make more validly than less-trained personnel, then the assessment process can be analyzed segmentally to see where and how this increased accuracy comes about. This would involve the same considerations mentioned above, with one important difference: now the clinician would be restricted to one specific instrument or one assessment technique. At this stage one finds many studies aimed at validating a specific technique, yet very rarely do they attempt to approximate the criteria already considered. For this reason, the question of what each diagnostic technique contributes to over-all diagnostic competence lies still unanswered.[3]

The present paper does not concern itself with the value implications of what the clinician does. Only in the sense that his work is compared with less-trained personnel is any judgment made of the worth of his labors. Whether diagnostic work-ups utilize the clinician's time to best advantage, whether he should be able to predict overt behavior, whether he should concentrate on research and/or therapy—all such questions are omitted here. The assumption has been made that clinicians spend a good deal of their time in diagnostic testing, and the question asked here is, "how good a job are they doing?"

---

[1] The author wishes to express his gratitude to the personnel of the Ann Arbor VA Hospital for their help in making this study possible. Special thanks are due Philip A. Smith of the hospital and E. Lowell Kelly and Max Hutt of the University of Michigan for their encouragement and criticisms of this paper.

[2] Now at Stanford University.

[3] These comments circumvent two related problems: (*a*) the "clinical-actuarial" controversy and (*b*) the "molar-molecular" controversy. Both issues lend themselves to parallel research within the framework mentioned above. For example, diagnoses arrived at actuarially could be compared with those of the clinician and the nonpsychologist. And, diagnoses made, using all the clinician's techniques, could be compared with ones made using a portion of these techniques or just one "favorite" technique.

The experimental answer to this last question hinges on two ancillary ones: who is to be called a clinician, and what is the nature of the independent evidence to test his judgments. The frequent criticism heard when groups of clinicians "fail" in some experimental study is that they were not "expert" enough or that they were essentially "academicians." Moreover, the criteria typically utilized in studies of this kind are all too often dubious in nature. Even the broad classifications "psychotic," "neurotic," "character disorder," and "normal"—much less the more homogeneous nosological categories such as "paranoid schizophrenia," "obsessive-compulsive neurosis," etc.—have no commonly accepted operational definitions. For this reason, it is typical to accept as a criterion either the consensus of many (the majority of whom *could* be in error) or the judgment of an established few (typically psychiatrists, who are usually responsible for the diagnostic referral to the clinical psychologist in the first place).

In this paper, which reports the first of a proposed series of assessments of clinical practice, the major concern is with diagnosis rather than prediction to later behavior. The nosological category chosen for this study was "organic brain damage, cortical," because of its inherent criterion for ultimate operational definition; namely, the independent diagnosis of a competent neurological team.[4] The clinicians assessed included those who were currently practicing as diagnosticians (either staff or trainees) at a large VA hospital, and their diagnostic performance was compared with that of nonpsychologists (hospital secretaries). This study was aimed at the second stage of a total assessment project, an appraisal of the validity of a specific diagnostic instrument in the hands of practicing clinicians. Since the Bender Visual-Motor Gestalt Test is the most widely used test for organic brain damage at the installation under consideration (and perhaps at many other installations), this instrument was chosen as the technique in question.

Procedure

Protocols of Bender-Gestalt tests were randomly selected from the files of a VA general medical and surgical hospital. Of these protocols, those from the first 15 patients who had been diagnosed by independent neurological examination as showing clear-cut evidence of cortical impairment were selected to represent patients manifesting organic brain damage (hereafter termed organics). As the nonorganic control group (hereafter termed nonorganics) the protocols of the first 15 patients from psychiatric wards were selected where (*a*) psychiatric diagnoses were clearly agreed upon, (*b*) no symptoms usually associated with organic brain damage were reported, (*c*) no record of cerebral trauma was noted, and (*d*) any routine examination by the neurological staff was negative to cortical impairment. These latter psychiatric patients were all fairly recent admissions to the hospital at the time they were tested and in general could be characterized as displaying acute rather than chronic symptomatology. Table 1 summarizes certain descriptive variables for these 30 patients.

The 30 Bender protocols were divided into three groups of 10 each, such that in Group I there were 2 organics and 8 nonorganics, in Group II there were 5 of each, and in Group III there were 8 organics and 2 nonorganics. This was done so as to be able to investigate the relationship between frequency of occurrence of a diagnostic entity (base-rate) and accuracy of its diagnosis; also, this allowed the work for each clinician to be broken down into more convenient sections. All protocols were assigned a code number which was

---

[4] This statement is not meant to imply that present neurological techniques are as yet able to assess the many gradations along the continuum from complete absence of cortex to perfect cortical health, at least short of opening up the skull and examining the brain microscopically. Nor is there any implication that localization, chronicity, and amount of cortical destruction will not influence performance on psychological tests. However, if a neurological team can separate cases on both extremes of the continuum, then these cases might serve as logical criteria for the evaluation of psychological techniques.

Table 1

Description of Patient Populations

Organics (N = 15; Mean age = 38; Age range = 24–61)						Nonorganics (N = 15; Mean age = 32; Age range = 23–54)	
Etiology (where available)	N	Localization (where available)	N	Symptomatology (where available)	N	Final Psychiatric Diagnoses	N
Trauma	7	Rt. cerebral hemisphere	1	Convulsions	4	Paranoid schizophrenia	4
Tumor	3	Lt. temporal lobe	1	Death	3	Catatonic schizophrenia	1
Multiple sclerosis	2	Rt. parietal lobe	1	Chronic brain syndrome	1	Manic-depressive psychosis	1
Thrombosis	1	Rt. front parietal lobe	1	Hemiparesis	1	Character disorder	4
Alvheimer's disease	1	Rt. carotid artery	1	Hemiplegia, spastic	1	Conversion reaction	2
				Headaches	1	Anxiety neurosis	2
						Obsessive-compulsive neurosis	1

printed on cardboard and stapled over the patient's name, thus removing this identifying data. On the other hand, all descriptions of how the patient actually drew the Bender designs (i.e., arrows, circles, short notes, etc.) were left on the protocols.

Only the actual reproductions of the Bender designs were used in this study, since uniform elaborations and free associations were not available for all the patients.

Table 2 gives a breakdown of some variables concerning the judges who participated in this project. All of the psychologists were actively engaged in diagnostic evaluations at the time the study was conducted, although they varied both in their general diagnostic testing experience, as well as in their particular experience with the Bender. Many used the test almost routinely as part of their psychological examinations of referred patients, while a few used it only rarely. Without exception, the nonprofessional judges had had no contact with the technique.

Each judge was given the three packets of 10 protocols, one packet at a time in random order. Directions were essentially as follows for all participants:

> You will be given 30 Bender protocols for your diagnostic impressions. For your convenience they have been divided into 3 groups of 10 each. Please judge each Bender individually, using any system you normally apply to such a task. Take all the time you want, and feel free to utilize any instruments (i.e., compass, protractor, ruler, etc.) which you feel will increase your diagnostic accuracy. Please do the very best job you possibly can.[5] Record your judgments on the face-sheet attached to each packet.

This face-sheet was a mimeographed form, listing the patients' code numbers in a column down the left-hand side of the sheet. Along the top of the sheet were column headings Organic and Nonorganic, as well as a confi-

Table 2

Description of Participating Judges

Group	N	Mean Age	Level of Training in Psychology	Approximate Experience with the Bender	
				Mean	Range
Psychology staff	4	35	Ph.D. plus 4–10 yrs. experience	6 yrs.	4–9 yrs.
Psychology trainees	10	27	M.A. plus 1–4 yrs. experience	3 yrs.	1–4 yrs.
Nonpsychologists	8	25	None	0	None

---

[5] To increase the judges' involvement in this diagnostic task, a bottle of Scotch was offered to the judge who performed most accurately. It was generally felt that all of the judges "tried their best." Although the judges spent only 15 to 30 minutes diagnosing all 30 patients, in general they expressed their impressions that they had been as careful in these diagnoses as they would be typically in their regular professional evaluations.

Table 3

Diagnostic Accuracy by Groups of Judges

Group	N	Mean % Correct	Range	Number of Judges Differing Significantly from Chance (50%) at the .05 Level
Psychology staff	4	65%[a]	60–70%	1 (25%)
Psychology trainees	10	70%[a]	60–77%	6 (60%)
Nonpsychologists	8	67%[a]	57–73%	5 (62.5%)
All groups	22	68%	57–77%	12 (54.5%)

[a] All mean differences between groups are nonsignificant.

dence rating scale labeled Positively; Fairly certain; Think so; Maybe; and Blind guess. Instructions on the form were to check the most appropriate diagnostic description for each patient and also to indicate one's degree of confidence in each of these judgments. Thus, each participant made one dichotomous diagnostic judgment and then qualified it on a five-point scale of confidence.

For the nonprofessional participants, the directions were amplified slightly. Since they had no basis at all for their judgments, they were administered the Bender as a patient would have received it. Then, they were given the following directions:

> Now you have an idea of the test itself. Psychologists use this test to help them diagnose patients with brain damage, on the assumption that brain-damaged persons draw the designs differently than persons without such damage. If you wish, you may use your drawings as a guide to the way a non-brain-damaged person responds on the test.

Then followed the same directions given the psychologists.

After the 22 judges had completed the diagnostic task, the Bender reproductions were scored by the Pascal-Suttell Objective Index (Pascal & Suttell, 1951), for comparison purposes with the clinical judgments.[6]

Results

Interjudge Analysis

The most striking finding of this study is the complete overlap between groups of judges on diagnostic accuracy (see Table 3). Staff psychologists, psychology trainees, and nonprofessional persons did not differ from each other in their ability to differentiate organic from nonorganic patients by means of their Bender protocols. The judges' degree of successful diagnoses ranged from 57% to 77%, with six judges differing at the .01 level and six more at the .05 level from statistical chance (50%). The remaining ten judges did not differ in their diagnostic accuracy from that attributable to chance alone.

While the Pascal-Suttell Objective Index (1951) was developed to help differentiate psychotic from neurotic populations, it has been used for the diagnosis of organics (Bowland & Deabler, 1956). Cutting scores for this use of the index vary, and extensive normative data is as yet unavailable. If a Z score of 100 is used to separate the groups, the index accurately diagnoses 63% of the patients in this study (Fisher's Exact: $p = .04$). When the cutting score is lowered to 90, the percentage of successful diagnoses increases to 67% ($p = .01$). The optimum cutting score for this population seems to be around 80, at which point the index diagnoses with 80% accuracy ($p < .005$). As the cutting point is lowered below this, accuracy slowly falls away (77%, 73%, and 70% for Z cutting points of 70, 60, and 50, respectively); at a cutting score of 40 or below, only chance results occur.

To check on group differences in interjudge agreement, the percentage of agreement for each group of judges was computed for each patient. The *average* percentage of agreement over all 30 patients, for each group of judges, ranged from 78% to 85% (chance would predict 50%), with no statistically significant intergroup differences.

Although the groups appeared quite similar with respect to diagnostic accuracy and interjudge agreement, there were large differences between the groups in the amount of confidence they placed in their judgments (Kruskal-Wallis H test: $p < .01$). The nonprofessional judges—with no training or experience in Bender interpretation (and therefore no apparent reason for developing confidence in the technique)—were, as a group, much *more* confident in their judgments than were either the staff or the trainees. And, the trainees displayed more confidence in their diagnoses than the staff. The results present the surprising paradox of an inverse relationship between the amount of experience with the Bender and the degree of confidence placed in diagnoses made from it. Moreover, there was no relationship between individual diagnostic accuracy and degree of confidence!

---

[6] The author wishes to thank Ronald Ribler of Michigan State University for his help in scoring the Bender protocols.

To test whether the reduced confidence exhibited by the more sophisticated judges was the result of their finer discrimination between the easier and the more difficult judgments (with the more difficult ones being rated with less confidence), the average degree of confidence for each judge was computed for those cases he diagnosed correctly and again for those cases when he misdiagnosed the patient. The difference between these two averages (hereafter termed the index of discrimination), when computed for each group of judges, was found to show no significant intergroup differences. In general, judges were about as confident on cases they misdiagnosed as they were on those they diagnosed correctly. When the index of discrimination was used as the basis for ranking the judges, there was found to be no relationship between this measure and either total degree of confidence or total diagnostic accuracy.

A further measure was developed as an index of clinical judgment,[7] under the assumption that a misdiagnosis which had been given with little confidence should not count as heavily against the clinician as one which was given with great confidence. An index was constructed in the following manner: for each correct diagnosis qualified by more than his median degree of confidence, each judge was given two points; for each correct diagnosis given with less than his median degree of confidence, one point; for each misdiagnosis with greater than median confidence, minus two points; for each misdiagnosis qualified by less than median confidence, minus one point. When total scores were computed for each judge, analysis showed no differences between the psychologists, trainees, and the nonprofessional judges on this measure of clinical judgment.

An analysis was made of the judges' tendencies to over-call or under-call organicity in the population they were diagnosing. The trainees as a group made the most "organic" diagnoses, 13.5 per judge, and they were the nearest to judging the actual number of organics in the sample. The nonprofessionals made the least such diagnoses, only 8.5 per judge. These differences almost approach statistical significance (Kruskal-Wallis H: $p < .10$), but there was *no* relationship between the number of patients called organic and diagnostic accuracy. A measure of the dis-

Table 4

Performance of the Three Groups of Judges on Some Selected Indices

Group	Interjudge Agreement, Average %	Degree of Confidence[a, b]	Index of Discrimination[c]	Number called Organic[d]	Index of Clinical Judgment[e]
Psychology staff	78%	2.1	1.3	12.1	16.2
Psychology trainees	85%	2.5	1.1	13.5	20.6
Nonpsychologists	85%	3.1	1.1	8.5	18.8
Totals	84%	2.6	1.1	11.1	19.1

[a] Intergroup differences in degree of confidence significant at the .01 level (Kruskal-Wallis H test); all other differences nonsignificant.

[b] Scored 0 for "blind guess"; 1 for "maybe"; 2 for "think so"; 3 for "fairly certain"; and 4 for "positive."

[c] Scored by subtracting the confidence score on misdiagnoses from that on correct diagnoses.

[d] The actual number of organics was 15.

[e] For method of scoring see text.

---

[7] Suggested by Philip A. Smith of the Ann Arbor VA Hospital.

Table 5

Group Performance as a Function of Differing Organic Base-Rates[a]

Group	Diagnostic Accuracy	Interjudge Agreement	Confidence	Discrimination	Number called Organic	Index of Clinical Judgment
Group I (2 organic; 8 nonorganic)						
Psychology staff	72%	78%	2.2	.2	3.0	7.0
Psychology trainees	62%	84%	2.3	.6	3.7[c]	4.9
Nonpsychologists	78%[b]	88%	3.1	.1	2.0[d]	9.5
Total	70%	84%	2.6[e]	.3	2.6	7.0
Group II (5 of each)						
Psychology staff	57%	78%	2.1	.5	3.8	3.2
Psychology trainees	73%	81%	2.5	−.2	4.3[c]	7.2
Nonpsychologists	68%[b]	80%	3.1	.2	3.0[d]	5.9
Total	68%	80%	2.6[e]	.1	3.7	6.0
Group III (8 organic; 2 nonorganic)						
Psychology staff	65%	80%	2.0	.7	5.5	6.0
Psychology trainees	75%	89%	2.6	.1	5.5[c]	8.5
Nonpsychologists	56%[b]	89%	3.1	.2	3.5[d]	3.4
Total	66%	87%	2.7[e]	.2	4.8	6.2

[a] For explanation of indices, see Table 4 or text.

[b] Differences among nonpsychologists at the three base-rates significant at the .05 level (Friedman two-way analysis of variance).

[c] Differences significant at the .01 level (Friedman).

[d] Differences significant at the .05 level (Friedman).

[e] Differences between groups significant at the .05 level for each base rate (Kruskal-Wallis one-way analysis of variance).

Note.—All other differences nonsignificant.

crepancy between the number of organics called and the actual number in the sample was computed for each group; neither this measure nor the gross number of organics called was related to *any* of the previously reported indices. Table 4 summarizes the findings on these various measures.

Since the original material presented to the judges consisted of three packets, each containing different percentages of organic patients, the data could be reanalyzed to investigate base-rate differences on the variables already considered. In this respect, each packet can be thought of as a separate study, each comparable to one conducted at a different type of installation (for example, a GMS hospital, an NP hospital, a home for mentally deficients, etc.). All of the previously reported indices (accuracy, interjudge agreement, degree of confidence, discrimination, clinical judgment, and number of patients called organic) were computed for each group of judges for each of the three different organic base-rates. Table 5 summarizes these results.

The Friedman two-way analysis of variance (nonparametric) was run to test differences attributable to the differing base-rates. The only statistically significant differences uncovered were on the "diagnostic accuracy" and "number called organic" indices. The nonprofessionals were significantly more accurate in their diagnoses as the number of organics in the sample decreased (probably because of their consistent tendency to be the most sparing in their use of the diagnosis "organic"). All groups called more patients organic as the actual base-rates of organics in the sample increased (although in the case of the staff judges, the small size of their group prevented statistical significance).

The Kruskal-Wallis one-way analysis of variance (nonparametric) was carried out to test if any of the differences between mean group scores, at each base-rate, were statistically significant. Just as in the case of the over-all analysis, the three groups of judges were found to differ significantly *only* in their degree of confidence.

Interpatient Analysis

Sometimes the "majority opinion" of a group is more accurate than any of the opinions of its individual members. To check on this, a "group" diagnosis, generated by combining the diagnoses made by all of the 22 judges, was examined for each patient. The amount of agreement between judges ranged from complete agreement on one patient (a manic-depressive psychotic, correctly diagnosed by everyone as nonorganic) to a complete 11–11 split in opinion on two patients (both organics). There was less than 75% agreement on 8 of the 30 patients.

For those 28 patients on whom there were "majority" diagnoses, the group as a whole misdiagnosed two nonorganics (both paranoid schizophrenics) and five organics (of whom three had symptoms which included grand mal convulsions, one was a case of posttraumatic encephalopathy, and one had a deadly glioblastoma multiforme). Thus, if one counts the evenly divided cases as errors (since no group diagnosis was generated), the group as a whole correctly diagnosed 70% of the patients, a figure not significantly different from the mean of the individual diagnoses (68%).

Interestingly, the degree of agreement among the judges did not correlate significantly with the *accuracy* of their combined diagnoses. Moreover, majority accuracy did not turn out to be related to such variables as the chronological age of the patient nor to his nosological category. On the other hand, since most of the judges tended to underestimate the actual number of organics in the sample, they tended to be most in agreement on patients whom they diagnosed nonorganic (Fisher's Exact: $p = .013$).

Certain patients must have seemed easier to diagnose than others, since there was a strong relationship between the amount of agreement on a patient's diagnosis and the total pooled confidence ratings given for this diagnosis (Fisher's Exact: $p < .001$). Surprisingly, however, the judges were just as confident in rating their incorrect as their correct diagnoses; nor were they any less confident in diagnosing organic than nonorganic patients.

A Subsequent Exploration

In the course of examining the relationship between diagnostic accuracy and experience with the Bender, it was noted that the judge who performed most accurately was a trainee who had spent considerable time (as part of his research for a doctoral dissertation) in administering, scoring, and interpreting tests for organic brain damage with a large group of brain-damaged patients. Although a staff judge also had considerable Bender experience and only performed at the median level on this task, one might still wonder whether specific intensive experience with the instrument might not increase diagnostic accuracy. In effect, this hypothesis would imply that although the practicing diagnosticians were not more accurate than nontrained persons on this task, real "experts" with the Bender could surpass them all.

To test this hypothesis, one of the country's foremost authorities on the Bender test was solicited to take part in this study.[8] This judge, taking some 20 hours to complete the diagnostic process, did perform more accurately than anyone in the original study—diagnosing 83% of the patients correctly. His scores fell in the middle of the over-all distribution on degree of confidence and discrimination, but he was one of the most accurate in judging the actual number of organics in the sample. He also was at the top on the index of clinical judgment.

Since his performance lends support to the "expert" theory on Bender diagnosis, it seemed legitimate to combine the scores of the top two diagnostic judges in this study into a subgroup of Bender "specialists" and then to reanalyze the data comparing their performance on all the indices with that of the three other groups. This analysis revealed no significant differences on measures other than those of accuracy and clinical judgment (Fisher's Exact: $p < .004$), on which variables the "specialists" were, in effect, selected.

---

[8] Special acknowledgment is due Max Hutt of the University of Michigan for the time and thought he invested in this phase of the study.

Discussion

In general, the results indicate that diagnostic accuracy (when using the Bender to diagnose organic brain damage) does *not* depend on experience or training in psychology (unless, perhaps, that training includes years of intensive work with the instrument in question). If he is not a real expert in the use of the Bender, a clinician will find that his secretary can probably do this particular job of differential diagnosis as well as himself. Moreover, she will most likely have considerably more confidence in *her* judgments than he would have in his. This makes it all the more unfortunate that one's degree of confidence bears no relationship to his diagnostic accuracy on this task!

Now these results, in themselves, may be embarrassing. And, when one considers the base-rates of organics usually encountered in clinical practice, one might even become alarmed. For in most settings the actual base-rate is closer to Group I (20% organics) than to any of the other groups in this study. And it is in this Group I where the nonprofessionals—by virtue of their tendency to label most patients nonorganic—appear most likely to overshadow their professional employers in diagnostic accuracy (see Table 5). However, neither the nonprofessionals nor the group as a whole did as well as could have been done by merely calling all patients in this group "nonorganic" and thereby diagnosing with 80% accuracy.

Before judging the Bender too harshly, however, the following factors should be considered: (*a*) the group as a whole did perform significantly above chance (50%) on this task, thus supporting the premise that *groups* of organics do respond differently to the Bender test than do *groups* of nonorganics (but when discriminable differences do appear they are typically so obvious that almost everyone can detect them); (*b*) this study provided no basis for comparing the Bender with other tests for organic brain damage in order to see whether the widespread faith in this technique is indeed comparatively justified; (*c*) it is of course possible that other tests taken in combination with the Bender may permit judges to diagnose more accurately; and (*d*) one cannot immediately discount the utility of the cues furnished by a face-to-face encounter with the patient, cues which were totally absent in this study.

On the other hand, these results might have been anticipated on the basis of previous work in this area (Bowland & Deabler, 1956). For example, Pascal and Suttell (1951), while formulating their Objective Index to differentiate psychotics from neurotics, have this to say about the differential diagnosis of organics:

> The Bender-Gestalt test cannot, in the absence of other data, answer that question (is there cortical damage?) except occasionally in extreme cases which are also clinically apparent (p. 40).

> Performance on the Bender-Gestalt test can indicate damage to the cortex only when the damage shows its effect by pronounced disturbance of the ability to execute the test. We know that nine year old children can reproduce the designs without marked deviation from the stimuli. When, therefore, an individual is functioning at a maturational level of nine years with respect to his ability to reproduce the designs, so to speak, we cannot distinguish between his deviation and those of individuals suffering from psychogenic disorders. This fact suggests that damage to the cortex has to be rather severe in its effect on the functioning efficiency of an adult of normal I.Q. before it can be detected by means of performance on the Bender-Gestalt test. This fact also suggests that actual lesions may exist which cannot, on the basis of the deviations noted by us, be detected in performance on this test (pp. 62–66).

Nevertheless, since the evidence suggests the possibility that real experts in the technique may perform with increased diagnostic accuracy on this task, it is conceivable that they might be able to communicate whatever interpretive refinements they possess. In effect, this has been tried with the Pascal-Suttell Objective Index, which at its optimum cutting point performed about as well as did the best of the individual judges. Significant in this connection, however, is the considerably greater length of time taken by the top expert in making his diagnoses as compared to the amount taken by any of the others, and correspondingly the greater length of time needed to score records by the Pascal-Suttell method. Assuming that real experts—after considerable time and careful scrutiny—can slightly out-perform hospital secretaries, the value of this potential increment in accuracy must be carefully evaluated.

Summary

Staff psychologists, psychology trainees, and nonprofessional (secretarial) judges made diagnostic judgments from the Bender protocols of 15 organic and 15 nonorganic patients and then indicated their degree of confidence for each of their diagnoses. The three groups of judges did not differ in their ability to diagnose organic brain damage from the Bender, although the nonprofessionals displayed considerably more confidence in their judgments than did either of the other groups. The Pascal-Suttell Objective Index approximately equaled the clinical judgments in diagnostic accuracy, but a renowned Bender expert was able to better the diagnoses of the practicing clinicians. The group as a whole diagnosed above a chance level, but when the base-rate of organic patients typically encountered in clinical practice is considered, the results suggest that chances for misdiagnosis could be increased by utilizing the Bender-Gestalt test.

Received January 14, 1958.

REFERENCES

BARRON, F. Some test correlates of response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 235–241.

BOWLAND, J. A., & DEABLER, H. L. A Bender-Gestalt diagnostic validity study. *J. clin. Psychol.*, 1956, **12**, 82–84.

HOLTZMAN, W. H., & SELLS, S. B. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.*, 1954, **49**, 485–490.

KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. Michigan Press, 1951.

PASCAL, G. R., & SUTTELL, BARBARA J. *The Bender-Gestalt Test*. New York: Grune & Stratton, 1951.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

CHANGE IN MENTAL ABILITY AS A FUNCTION OF TEST ANXIETY AND TYPE OF MENTAL TEST[1]

FREDERICK LIGHTHALL, BRITTON RUEBUSH, SEYMOUR SARASON
Yale University

AND IRVING ZWEIBELSON
Milford, Connecticut Public Schools

Zweibelson (1956) reported the results of an investigation of the relation between test anxiety in fourth-grade children and their performance on test-like and game-like group intelligence tests. The three group tests of mental ability employed in his study, the Otis Alpha, Otis Beta, and Davis-Eells tests, were described in detail with regard to items, style of writing, method of administration, and the like. Stronger relationship was found between the Otis Alpha test (a test-like instrument) and the Test Anxiety Scale for Children (TASC)[2]—$r = -.28$—than between the Davis-Eells test (a game-like instrument) and the TASC—$r = -.14$. While the correlation between the Beta test and the TASC ($-.24$) was not significantly different from the correlation between the Davis-Eells test and the TASC ($-.14$), the Beta-TASC correlation was similar in magnitude and direction to the Alpha-TASC correlation. These correlations suggest that the two Otis tests together are different from the Davis-Eells test relative to performance on the TASC. Although the correlation coefficients themselves are not large, the correlation differences raise the question whether level of test anxiety is a function of the *type of test* being administered. Davis and Eells, in their manual (Davis & Eells, 1956, pp. 4–5), contend that performance on their test is less a function of traits like competitiveness, compulsiveness, and anxiety than is performance on other tests which emphasize reading ability and speed. Zweibelson found some support for their contention in his study in which fourth-grade pupils were administered the two Otis tests, the Davis-Eells Games, and the TASC once only. The present paper throws further light on the relationship between test anxiety and group means on these three tests of mental ability. In addition, data are presented which show the relationship between test anxiety and *changes* in Otis Beta and Davis-Eells scores over an interval of two years.

Our specific interests can be stated in the form of three hypotheses: First, the difference between low anxious (LA) and high anxious (HA) group means should be greater on the Otis tests than on the Davis-Eells tests, the HA group obtaining lower means. Second, the increment in mean raw score from one year to the next should be greater for the LA than for the HA groups on the Otis tests. And third, the difference in increment between the LA and HA groups should be less on the Davis-Eells tests than on the Otis tests. Since data for this comparison were available for only the Otis Beta test, the Otis Alpha scores could not be compared with those of the Davis-Eells.

All three hypotheses stated above follow from essentially the same rationale regarding test anxiety and test performance. Sarason, Davidson, Lighthall, and Waite (1958, p.

---

[1] This research was supported by a grant from the United States Public Health Service.

[2] This scale is described in detail (Sarason et al., 1958). The TASC has been shortened, subsequent to its 1954 administration, from the original length of 43 items to its present length of 30 items. The form employed in the analysis presented in this paper is the original form.

109) conceptualize test anxiety as "primarily interfering in its effect." Zweibelson (1956) qualified this conception. He interpreted his data as indicating that the test-anxious reaction was more interfering in performance on a test-like than on a game-like task. In exploring available data with regard to the three hypotheses stated above we sought further to extend the concept of test anxiety. For example we asked, in reference to Hypothesis 2, "Does the differential performance of high and low test-anxious subjects to different types of mental test (test- and game-like) hold up also for mean *change* over time on these same types of test?" Our answer as stated in the second hypothesis was again a reflection of our assumption that test anxiety is interfering in its effect, but differentially interfering with performance on test-like and game-like tests. In testing this hypothesis we were testing whether the test-anxious reaction was interfering with *improvement* in performance on these two types of mental test. The first hypothesis, in contrast, is simpler in its assumptions and follows more directly from the implications of Zweibelson's study, namely, that over-all mean performance on these two types of test should be differential as regards the high and low anxious groups, the LA group doing significantly better than the HA on the test-like test but not so markedly superior to the HA on the game-like task, where anxiety is assumed to be less interfering. Hypothesis 3, like the second hypothesis, relates to change in performance on the two types of test and indicates our assumption that this change should be differential as regards the HA and LA groups. That is, the scores of the HA should increase significantly less than those of the LA on the test-like instrument, whereas the gains of the LA and HA should not be significantly different from each other. Each of these hypotheses assumes that the TASC is effective in discriminating the high from the low anxious over a period of time up to two and a half years from the date of TASC administration.

Table 1

Sequence of Test Administrations

	Date	Grade	Date	Grade	Date	Grade
Otis Alpha ($N^a = 108$)	Jan., 1952	2			May, 1954	4
Otis Beta ($N^a = 92$)	Oct., 1954	5	Oct., 1955	6	Oct., 1956	7
Davis-Eeels ($N^a = 92$)	Oct., 1954	5			Oct., 1956	7

[a] N here refers to the number of *S*s finally selected for study from the total tested population. While these *S*s cannot be considered to be a purely random sample, it is safe to say that the only factor limiting their number was absence from school on testing days.

Subjects

Subjects were all in the same grade. Ninety-two pupils were distributed equally in four sex-anxiety groups ($n = 23$) in all but the first of three analyses. In that first analysis 108 pupils, including 83 of the above-mentioned 92, were distributed equally between the high and low anxiety groups ($n = 54$), but unequally between the two sex groups, boys numbering 62 (31 LA and 31 HA) and girls numbering 46 (23 LA and 23 HA). Data were available for these *S*s because of their connection, as a part of a larger sample, with a continuing research program. These data are not the result of a planned execution of a research design but, rather, result from an investigation after the fact, so to speak, to establish whether certain effects found by Zweibelson (1956) are also present in these data, resulting from the regular school testing program.

Procedure

The school-wide testing program had been carried out in the sequence shown in Table 1. The TASC was administered to these *S*s in May, 1954, when they were in the fourth grade. Unlike Zweibelson's study, the LA and HA groups in the present study were not taken from the extremes of the distribution of test anxiety scores but were grouped relative to their position above or below the sample median. Separate medians were computed for boys and girls. Analysis of variance designs were used throughout, with both between and within subject effects. In comparing scores of two different tests standard scores were computed. Analysis of gain scores in the Otis Beta-Davis-Eells comparison was based on the standardized differences of each individual's raw score on the first administration of a test subtracted from his raw score on that test administered two years later.

Results

The first analysis, involving the TASC and the Otis Alpha test, showed no anxiety, sex, or anxiety-sex interaction effects either on over-all Otis Alpha means of the two administrations combined or on gain in mean from the first to the second administration.

The second analysis, involving the TASC and the Otis Beta test and in which there was a 77% overlap with *S*s employed in the first analysis, showed the only clearly significant finding ($p < .01$) to be a simple main effect of sex grouping on the over-all Beta mean (collapsing the three sets of scores from the three Beta administrations into one set of scores). The girls received a higher over-all mean than the boys. In addition, this mean was different for the two anxiety groups between the .10 and .20 level, the HA doing poorer than the LA groups, as expected. There was also a difference ($.10 < p < .20$) in the rate of gain in Beta score over the three years for the two anxiety groups, the LA gaining the more rapidly and augmenting the difference between the HA and LA means. Sex was not found to affect the rate of gain on Beta score.

The third analysis, with the same *S*s as in the second, concerned the differential relationship of test anxiety to Davis-Eells scores compared with Otis Beta scores. This analysis revealed a significant effect of anxiety level on mean mental ability score ($p < .01$) which was uncomplicated by interaction of test (Otis Beta or Davis-Eells) with anxiety. As to the change of scores on the two tests over two years, there was no over-all effect of anxiety on gain but rather an anxiety-test interaction effect, as hypothesized ($.05 < p < .10$). This effect, while not robust, revealed that while the LA gained more over two years than HA on the Beta (test-like) test, the HA gained more than the LA on the Davis-Eells (game-like) test. The difference between LA and HA was more marked on the Beta test than on the Davis-Eells test, however. The third analysis, including mean squares for both over-all mean effects and for effects of change, is summarized in Table 2.

The reader will note that the within *S*s and between *S*s error mean squares for (standardized) gain scores are of virtually equal magnitude. This indicates that there is as much variation, on the average, from an individual's gain on the Otis to his gain on the Davis-Eells as there is from one individual's over-all gain to another's over-all gain. This is in contrast to the within and between error terms, also appearing in Table 2, in the analysis of raw scores combined over the two years. In other words, there is no correlation between *gains* on the Otis and *gains* on the Davis-Eells over this two-year period, even though there is a correlation between over-all raw scores on the two tests as indicated by the small magnitude of the within *S*s error term for raw scores (163.43) relative to the comparable error term between *S*s (450.05). Pooled within groups correlations between Otis and Davis-Eells scores from first and second administrations of the tests, respectively, are .46 and .59.

Table 2

Summary of Otis Beta and Davis-Eells Raw Score and Gain Score Analyses

($n = 23$, $N = 92$)

Source	*df*	Mean Square Raw Score	Mean Square Gain Score
Between	3		
Anxiety	1	3,213.91**	.24
Sex	1	364.57	.07
A×S	1	219.58	.00
Error	88	450.05	1.03
Within	4		
Test	1	16,209.39[a]	.00
T×A	1	66.98	3.30*
T×S	1	147.96	.47
T×A×S	1	151.56	.07
Error	88	163.43	1.01

* $.05 < p < .10$.
** $p < .01$.
[a] Significant but trivial, since standardized scores were not used and tests contained different numbers of items.

Discussion

The first hypothesis stands unconfirmed: over-all mean performance on the two Otis tests is not found to be related to test anxiety in any way different from that in which the over-all mean on the Davis-Eells test is related to test anxiety. The second hypothesis receives support from the analysis of Otis Beta increments over three years, but not from analysis of Otis Alpha increments over two (earlier) years. Alpha gains from the second to the fourth grade are not significantly related to (fourth grade) test anxiety grouping, while Beta gains from the fifth to the seventh grade are somewhat ($.10 < p < .20$) related to (fourth grade) test anxiety grouping. The third hypothesis is confirmed. Furthermore, not only is the *magnitude* of the difference between the gain scores of the HA and LA groups less on the Davis-Eells test, but, in comparison to the LA, the HA show *greater* improvement on the Davis-Eells and *less* improvement on the Otis Beta.

Various interpretations may be offered for the lack of correlation between gain on the Otis Beta and gain on the Davis-Eells tests. First, it is well known that error in gain scores is greater than error of measurement in either the first or second measurement taken separately. Thus, unreliability might appear to be suppressing the correlation between the two gain scores. A valid objection to this argument would seem to be that the gain scores were reliable enough to yield a within *S*s effect (test-anxiety interaction) significant between the .05 and .10 level and therefore should be reliable enough to allow for correlation with each other if such relation did in fact exist. Another explanation for the lack of relationship between gains on the two tests is that the sample employed might be peculiar in this regard. While it is true that the sample was not drawn randomly from an unselected group of pupils, there still remains the discrepancy between correlations, .46 and .59, of raw Beta and raw Devis-Eells scores, for first and second testings on the one hand, and the correlation between gain scores (zero), on the other hand. A third explanation for the lack of correlation, with severe implications for psychological measurement, is that improvement on these two tests, the Otis Beta and Davis-Eells, requires quite independent types of growth, even though they both purport to measure intelligence. If this is true, it lends strong support to Davis' and Eells' contention that they are tapping a different domain from those measured by most group-administered mental tests, of which the Otis is in many respects representative. If gains on group-administered mental ability tests are important, and if the results presented here on gains are at all representative, then the choice of which test and/or type of test (examination- or game-like) to employ is extremely important. Additional research on the reliability and correlates of the gain scores on the Otis Beta and Davis-Eells tests is necessary before unequivocal interpretation of longitudinal results with these tests is possible. The assumption that the TASC is effective in discriminating the high from the low test-anxious over a period of time up to two and one-half years appears to have valid foundation, provided that this period is between the fourth and seventh grades.

While the first hypothesis was not borne out and the second received only partial support, it must be emphasized that the basis of grouping *S*s in anxiety levels was their position relative to the median and not membership in extreme groups, as in the Zweibelson study. This fact tends to point up the significance of the support for the third hypothesis. However, as this was an exploratory study, its main significance lies in revealing general trends and providing suggestions for further research. Two general trends are shown: (*a*) mental ability scores tend to change from fifth to seventh grade differentially for high and low test anxiety groups and (*b*) this change is related to the type of test used in measuring mental ability. Further research suggested by these results concerns (*a*) the interaction of grade with the effect of anxiety on change in mental ability scores (e.g., why is there no effect of anxiety on change from the second to the fourth grade when there is one from the fifth to the seventh grade?) and (*b*) the specific nature of the effect of different types of mental ability test on the relation between anxiety and performance on these tests. For example, does the difference in HA and LA performance on the Otis Beta from that on the Davis-Eells test show that where no reading ability is needed (Davis-Eells) anxiety is not inhibitory and even possibly facilitating in effect on performance? The above-mentioned trends have at least two implications. First, the earlier the anxiety level of school children can be identified, the earlier they can be provided with a therapeutic program designed to help them recognize the interaction between their anxiety and various types of situations, and to learn to capitalize on the (possibly) facilitating effects of anxiety while minimizing its interfering effects. Second, the interpretation of change in mental ability must take into account variations in type of measuring instrument such as Zweibelson has enumerated (1956, Table 1).

Summary

Three analyses were carried out on data available from a regular school testing program. Each analysis concerned the relation of performance on the Test Anxiety Scale for Children to performance on three mental ability tests described and employed by Zweibelson (1956). The effect of test anxiety grouping (above and below median) on both over-all mean and change in mean over a duration of two years was tested with reference to three hypotheses: 1. The difference between low anxious (LA) and high anxious (HA) over-all means should be greater on the Otis (test-like) tests than on the Davis-Eells (game-like) test. 2. The mean increment from one year to the next should be greater for the LA than the HA on the Otis tests. 3. The difference in increment for LA and HA should be less on the Davis-Eells test than on the Otis Beta test. The first hypothesis was not supported, while the second received partial support. The third hypothesis was supported. Implications both for practice and further research were discussed.

Received January 18, 1958.

REFERENCES

Davis, A., & Eells, K. *Davis-Eells test of general intelligence or problem-solving ability.* Yonkers, N. Y.: World Book, 1953.

Sarason, S. B., Davidson, K., Lighthall, F., & Waite, R. A test anxiety scale for children. *Child Developm.*, 1958, **29**, 105–113.

Zweibelson, I. Test anxiety and intelligence test performance. *J. consult. Psychol.*, 1956, **20**, 479–481.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

THE VALIDITY OF THE SPIRAL AFTEREFFECT AS A CLINICAL TOOL FOR DIAGNOSIS OF ORGANIC BRAIN PATHOLOGY[1]

EMILY B. PHILBRICK
University of California, Berkeley

Many of the cases that are referred to the psychology staff in a general hospital are ones in which the neurologists wish additional information in their attempt to determine whether or not organic brain pathology exists. Although psychological tests such as the Bender-Gestalt, WAIS, and Rorschach are useful in such a diagnosis, these tests are far from adequate (Armitage, 1946; Yates, 1954). More clinical tools are vitally needed and so any procedure needs to be thoroughly investigated that reportedly will detect organic brain pathology.

Price and Deabler (1955) and Farrett, Price, and Deabler (1957) reported that 96% to 98% of cases with organic brain syndrome could be picked up by the Archimedes Spiral test, while Gallese (1956), with the same apparatus and a somewhat altered procedure and sample, reported that 66% of such cases could be correctly diagnosed. These reports appeared to be most encouraging and a cross validation study of the work of Price and Deabler was undertaken. In addition, the spiral test was run after giving most patients the Weinstein sodium amytal test (Weinstein & Malitz, 1954). This Weinstein test presumably increases certain organic symptoms when diffuse brain pathology exists. Therefore, it seemed important to study its effect on the perception of the spiral aftereffect.

Method

Eighty-one patients, essentially consecutive admissions, from the neurology ward of this general hospital were given the Archimedes spiral aftereffect test. Only those physically unable to leave their beds, or with gross eye defects, were omitted from the testing. The apparatus and procedure used corresponded to that reported by Price and Deabler (1955).

An electric motor commonly used for color mixing experiments was used to rotate a white 8″ disk on which was painted a black Archimedes spiral of 920° or $2\frac{1}{2}$ circuits about the center. The apparatus was so constructed that the disk could be rotated in either direction. The disk was rotated at approximately 100 rpm.

The patient was seated 8 ft. from the apparatus. Testing was conducted at noon when natural illumination was good. The patient was informed that this was a special eye test. He was asked to keep his eyes on the center of the spiral and was reminded of this at intervals during the testing period. In half of the cases the spiral was rotated first to give a negative aftereffect of expansion (Spiral A), while in the other half the rotation was in the opposite direction first, giving a negative aftereffect of contraction (Spiral B). After 10 seconds of rotation the patient was asked: "What does the line seem to be doing?" At the end of 30 seconds the rotation was stopped and the patient was asked immediately: "Now what does the line seem to be doing?" The patient's answer was recorded verbatim. Each patient was shown the spiral four times, in the order A-B-A-B or B-A-B-A.

---

[1] This research was carried out at the Veterans Administration Hospital, San Francisco, Calif., upon the suggestion of Jerome Fisher. The author wishes also to thank Stuart MacRobbie, who administered the sodium amytal injections and did the neurologic classification, and Henry W. Neuman and Lewis A. Roberts, who also aided in the neurologic classification. David A. Rodgers helped by reading critically the first draft of this manuscript. The manuscript was prepared while the author held a national fellowship awarded by the American Association of University Women.

The author is now at College of Marin, Kentfield, Calif.

The responses were scored in two different ways: once following the method outlined by Price and Deabler (1955) and Farrett et al. (1957), who used half scores as well as whole scores, and once following Gallese (1956), who used only whole scores. Price and Deabler (1955) state:

> Each normal report, that is, seeing the negative aftereffect correctly, was scored 1. Each abnormal report, that is, failure to perceive the negative aftereffect, was scored 0. Reports of aftereffect in which no apparent change of dimension occurred but in which apparent motion forward or backward was reported, were given a score of $\frac{1}{2}$. . . . The *S*s scored a total of 4, 3, 2, 1, or 0 depending on the degree of their perception of the aftereffect. . . . In the interests of conservatism, to give *S* the benefit of a doubtful response, fractional or half scores were raised to whole scores in final computation (p. 300).

Gallese (1956) does not distinguish between reports of change of dimension and apparent motion forwards or backwards and gives a full score of 1 to either type of response.

After the spiral test had been run, a Weinstein amytal test (Weinstein & Malitz, 1954) was run on 53 of the 81 patients, and this was then followed by another complete presentation of the spiral aftereffect test. The Weinstein test is designed to diagnose diffuse brain pathology and consists of asking the patient the same 21 questions before and after an injection of sodium amytal. Any changes in the direction of concretism, lack of orientation, etc. in the answers are noted. The hospital neurologist interested in checking this test was present throughout the spiral and amytal tests, and administered the drug. The amytal was given during a 10-minute period with the amount given varying in each case. The drug was administered until the patient showed signs of visual nystagmus, blurring of speech, drowsiness, and counting errors (backwards from 100 by ones).

The exact action of sodium amytal is not known. One theory states it may act to inhibit some enzyme system or cell metabolism, thus preventing the cells from functioning so rapidly. Unlike phenobarbital and other similar drugs, it not only affects higher centers but acts more specifically, and to a greater degree, on subcortical areas concerned with associations between cortical lobes and diencephalic centers. Because of this action, clinical manifestations of organic damage are augmented, and the spiral test was rerun as it seemed possible that the ability to see the negative spiral aftereffect might be further impaired when organic brain pathology was present.

At the time of working with these patients they had not yet been diagnosed as having an organic brain syndrome. It was not until all the testing had been completed that at least two neurologists rated each patient as "organic" or "nonorganic" on the basis of discharge summaries and total case material. A patient was placed in the "organic" category if he was considered to have organic pathology above the foramen magnum. On the other hand, if the neurologists considered the patient not to have organic pathology or to have organic pathology below the foramen magnum, he was placed in the "nonorganic" category. There were nine cases on which there was not agreement, and these were not considered as part of the final sample. Thus the results were computed on a sample of 72 cases, and on 53 cases in which amytal had been given. Forty-five cases of the 72 were considered to be organic. These included such disease entities as multiple sclerosis, brain tumors, Parkinsonism, epilepsy, cortical atrophy, etc.

Price and Deabler (1955) and Farrett et al. (1957) refer to their organics as having cortical involvement. Nowhere do they state exactly what they mean by this. Since their sample includes cases of Parkinsonism and cerebral vescular accidents, and these are commonly not considered as usually involving the cortex directly, it may be inferred that "cortical involvement" is used loosely to mean "brain involvement." Therefore, the sample of organics used in this study may be considered comparable to those in the former sample, with the addition of epileptics in this sample.

Results

As was previously stated, the data were analyzed using the scoring system reported by

Price and Deabler (1955) and Farrett et al. (1957) and that used by Gallese (1956). No rationale for using half scores is given by the former authors. Goldstein (1948) and other investigators consider that organic brain pathology leads to impairment of abstract ability, with a concreteness of thinking resulting. Possibly Price and Deabler are using this concept and considering the response of change of dimension as being more abstract than movement forward or backward and, hence, the superior response. It would seem that age, IQ, etc. might be important factors, rather than simply organicity, in determining the verbal response. Thus the validity of considering the form of response as diagnostic has been checked here by scoring the records with and without this differentiation. Price and Deabler are extremely vague in their description of their method of analyzing their data. Gallese considers scores of 0–2 as organic, and 3–4 as nonorganic. This, then, is the cutting point used here. Chi square was the statistic used to test the significance of the results.

As can easily be seen from the tables, these results differ greatly from those reported previously by Price and Deabler and Gallese.

Table 1 gives the results following the procedure and analysis used by Price and Deabler. Under the conditions of this study there is no differentiation of organics from nonorganics by the spiral aftereffect test. Several things probably account for these surprisingly discrepant results. Subjects who fall at extreme ends of a distribution as they do in the former studies (organics vs. normals) may often be correctly diagnosed, while the same test may fail with individuals falling closer to the center of the distribution. There are no normal cases in this study comparable to those used in the former work. However, as Fisher, Gonda, and Little (1955) and Stilson, Gynther, and Gertz (1957) have emphasized, the sample used here is the meaningful one for testing the efficacy of a clinical tool. Previously, the normal population had been drawn from hospital personnel. Here, all the patients had neurologic symptoms that had to be evaluated. Some later were considered on the basis of neurological tests (examination, EEG, pneumoencephalogram, etc.) to have organic brain pathology above the foramen magnum, and some were not. It is on such a sample that a clinical tool of this sort must prove its usefulness.

Table 1

Discrimination Value of Spiral Aftereffect Tests Using Price and Deabler's Scoring System (Before Amytal)

Spiral Criterion Price and Deabler (1/2 scores)	Neurological Classification: Organic	Nonorganic
Organic scores 0–2		
No. of patients	24	21
% of patients	53.3%	46.7%
Nonorganic scores 3–4		
No. of patients	13	14
% of patients	48.2%	51.8%

Note.—Chi square $p >. 70$.

Table 2

Discrimination Value of Spiral Aftereffect Tests Using Gallese Scoring System (Before Amytal)

Spiral Criterion Gallese Scoring (No 1/2 scores)	Neurological Classification: Organic	Nonorganic
Organic scores 0–2		
No. of patients	9	36
% of patients	20%	80%
Nonorganic scores 3–4		
No. of patients	4	23
% of patients	14.8%	85.2%

Note.—Chi square, $p > .55$.

Furthermore, at the time of testing, the final diagnosis of the patient was unknown. It is known this aids objectivity and lessens the danger of maximizing chance differences on the test between known groups. In the other studies, the investigators were working with known groups and could analyze their data with this in mind.

Half scores are not used in obtaining the results reported in Table 2. The results still do not show significance. Fewer organics are diagnosed correctly by this method than by the use of ½ scores. However, the number of false positives is considerably reduced. For this reason it seems preferable to use the scoring system of Gallese who does not employ ½ scores.

The results shown in Table 3 indicate that the amytal injection does make it more difficult to perceive the spiral aftereffect. Of the organics, 78.8% are correctly picked up. However, as there are 65.0% false positives, this procedure is useless as a diagnostic tool.

The results reported in Table 4 are the only significant ones. Half scores were not used here. Once again the indication is that the use of ½ scores is a poorer method of differentiation than is the use of whole scores. As in Table 2, there is a pick-up of 66.7% of the organics, with only 35.0% false positives. Although these results are significant (p .05), it is questionable whether the differentiation is clear-cut enough to warrant the patient's being subjected to this rather lengthy and disagreeable procedure at this time. Additional work is needed to determine the validity of this procedure and to give an indication of its usefulness as a clinical tool.

Under none of these conditions do the results in any way duplicate the work previously done with the spiral aftereffect test. Therefore, at the present time it seems that this test is not a useful clinical tool to determine the presence of organic brain pathology in a general hospital population.

Table 3

Discrimination Value of Spiral Aftereffect Tests Using Price and Deabler's Scoring System (After Amytal)

Spiral Criterion Price and Deabler (1/2 scores)	Neurological Classification	
	Organic	Nonorganic
Organic scores 0–2		
No. of patients	26	7
% of patients	78.8%	21.2%
Nonorganic scores 3–4		
No. of patients	13	7
% of patients	65.0%	35.0%

Note.—Chi square, $p > .25$.

Table 4

Discrimination Value of Spiral Aftereffect Tests Using Gallese Scoring System (After Amytal)

Spiral Criterion Gallese Scoring (No 1/2 scores)	Neurological Classification	
	Organic	Nonorganic
Organic scores 0–2		
No. of patients	22	11
% of patients	66.7%	35.0%
Nonorganic scores 3–4		
No. of patients	7	13
% of patients	35.0%	65.0%

Note.—Chi square, $p < .05$.

It was often noted that the organic patients who saw the aftereffect seemed to notice it for a shorter duration of time than did those without brain pathology. This was also noted by Gallese. A study is now in progress to determine whether the duration of the aftereffect will differentiate organics from nonorganics in this general hospital group. Although such procedures were not included in the present study, it might be valuable to study the effect of increasing the difficulty of seeing the aftereffect by reduced lighting, lowered speed of rotation, etc., rather than by the arduous process of administering sodium amytal.

Summary

1. Eighty-one patients admitted to the neurology ward of a Veterans Administration general hospital were given the Archimedes Spiral Aftereffect test. This was administered in the manner reported by Price and Deabler (1955) and Farrett et al. (1957), and scored according to their method and also by that of Gallese (1956).

2. Fifty-three patients received the Weinstein amytal test after being given the spiral test and were then retested on the spiral.

3. The diagnostic category into which the patients fell was unknown at the time of testing. After the sample had been run, at least two neurologists rated each patient as having, or not having, organic brain pathology above the foramen magnum. There was dis-

agreement in nine cases, so these were omitted from the analysis.

4. Analysis of the results indicates no significant differentiation, except after the amytal injection using Gallese's scoring system. However, the presence of 35% false positives with this rather involved procedure discourages its use as a clinical tool, at least until further research has purified the procedure and validated its usefulness. Thus it appears that in its present form the spiral aftereffect test is not useful for diagnosing organic brain pathology in a general hospital setting.

Received January 20, 1958.

REFERENCES

ARMITAGE, S. G. An analysis of certain psychological tests used for the evaluation of brain injury. *Psychol. Monogr.*, 1946, **60**, 1–48.

FARRETT, E. S., PRICE, A. C., & DEABLER, H. Z. Diagnostic testing for cortical brain impairment. *A.M.A. Arch. Neurol. Psychiat.*, 1957, **77**, 223–225.

FISHER, J., GONDA, T., & LITTLE, K. B. The Rorschach and central nervous system pathology: A cross validation study. *Amer. J. Psychiat.*, 1955, **111**, 487–492.

GALLESE, A. J. Spiral aftereffect as a test of organic brain damage. *J. clin. Psychol.*, 1956, **12**, 254–258.

GOLDSTEIN, K. *After-effects of brain injuries in war.* New York: Grune & Stratton, 1948.

PRICE, A. C., & DEABLER, H. Z. Diagnosis of organicity by means of spiral aftereffect. *J. consult. Psychol.*, 1955, **19**, 299–302.

STILSON, D. W., GYNTHER, M. D., & GERTZ, B. Base rate and the Archimedes spiral illusion. *J. consult. Psychol.*, 1957, **21**, 435–437.

WEINSTEIN, E. A., & MALITZ, S. Changes in symbolic expression with amytal sodium. *Amer. J. Psychiat.*, 1954, **111**, 198–206.

YATES, A. J. The validity of some psychological tests of brain damage. *Psychol. Bull.*, 1954, **51**, 359–379.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

THE YOUNG REBEL:
SELF-REGARD AND EGO-IDEAL [1]

EVA MARIA SHIPPEE-BLUM
San Mateo Community Hospital, California

This study was undertaken as a part of a larger investigation of the ego-structure of rebellious adolescents (Shippee, 1954). For this purpose, the psychoanalytical frame of reference was adopted since it provides a convenient conceptualization of underlying behavior determinants in terms of the construct of ego, varying on a continuum of ego-weakness to ego-strength. The primary hypothesis to be tested was that young rebels are endowed with less ego-strength than their more cooperative agemates. The deficiency in ego-strength was assumed to result in impaired functioning in reality adaptation for which an adequate ego is a prerequisite (Redl & Wineman, 1952).

Freudian theory accords the ego the role of mediator between the demands of successful survival in the world of reality and the demands of gratifying the inner world of instinctual impulses. Such a "reality orientation" depends upon an ego strong enough to tolerate unpleasant inner tensions until the time when instinctual needs can be appropriately satisfied without harm to the individual (or to others). Ego-strength refers to the ability to withstand tension without resorting to various defense mechanisms such as distortion, repression, displacement, uninhibited attack or flight. Such an ego should be able to make use of and enjoy opportunities maximally without undue anxiety, guilt, avoidance, etc.

The connection between "reality orientation" and "ego" is described by Fenichel (1954, p. 35): "The origin of the ego and the origin of the sense of reality are but two aspects of one developmental step."

The sense of reality, also designated as the "reality principle" in contra-distinction to the "pleasure principle," implies, as indicated above, the ability to wait, to postpone impulse gratification. A temporal delay is interpolated between the original stimulus-response unit typical for infant behavior. The long-circuiting of immediate reflex-type discharge phenomena is made possible by the emergence of intervening symbolic processes, i.e., the emergence of the ego. Fenichel (1954, p. 367 ff.) writes as follows:

> The infant as long as he acts according to the pleasure principle tries to discharge tensions immediately, and experiences any excitement as "trauma," which is answered by uncoordinated discharge movements. . . . They [impulse neurotics, i.e., delinquents, etc.] still act as if any tensions were a dangerous "trauma."

As mentioned before, in this study we are concerned with the ego's attribute of ego-strength. In order to pass from the level of abstraction to the behavioral aspects of ego-strength, Block's (1950) intervening variable "ego-control" is adopted in this investigation. "Ego-control" is to provide the transitional step from ego-strength as manifested in tension-binding-capacity,[2] to the experimental measure thereof. The ego-control system, according to Block, functions to maintain the most subjectively efficient balance for the individual between immediate gratification of needs and delay of gratification congruent with the requirements of the external world.

---

[1] This investigation was supported in part by a research grant (M-528) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. The study was carried out at Stanford University. Jeanne Block, Barclay Martin, and Joseph Luft collaborated in this research.

[2] Tension-binding-capacity as used here is the technical equivalent for the more descriptive terms: tension-tolerance, frustration-tolerance, impulse-control.

He states that the ratio—Immediate Gratification/Delay of Gratification (IG/DG)—tends toward maximization as a consequence of the hedonic postulate which most personality theories accept. Block terms individuals who more or less maintain an optimal IG/DG ratio as Adequate Controllers. This designation corresponds to what the present writer refers to as optimal ego-strength. Block distinguishes between the Adequate Controller and the Under Controller; the latter is one whose ego-control system has insufficiently developed the necessary tension tolerance. His ID/DG ratio is above the optimal range. We recognize in the individual with low tension tolerance one of the aspects of impaired ego-strength discussed previously as a factor in rebelliousness.

In this study the major hypothesis of decreased ego-strength in the rebellious adolescent was tested by comparing the tension-binding-capacity of a group of rebellious adolescents to an equated group of cooperative adolescents. As an operational definition of low tension-binding we adopted Block's (1950) inferred variable "under-control" and his IG/DG measure thereof which consisted of a questionnaire designed to assess under-control. The prediction which follows is: On an ego-control questionnaire designed to measure Impulse Gratification vs. Delay of Gratification, the rebel will be on the under-controlling end of the distribution of scores.

Two subsidiary hypotheses were derived from the primary concept of ego-weakness in the rebel. The first hypothesis was that the rebel's reality testing is inadequate. Specifically, it was predicted that one important aspect of reality testing is distorted, namely, the objective evaluation of the rebellious adolescent's own self. Indeed, a realistic self-evaluation is only possible if the individual is able to tolerate the frustration arising from an awareness of his limits. Such a frustration tolerance requires a strong ego. Consequently, a person lacking the prerequisite ego-strength, as exemplified by the rebellious adolescents, will show low frustration tolerance. Low frustration tolerance may manifest itself in various ways; for instance, in a narcissistic orientation toward the self.

Following Block (1950), it was posited that self-evaluation varies along a continuum ranging from unrealistically favorable to unrealistically unfavorable. Present evidence indicates that a realistic self-regard implies moderate self-esteem. A narcissistic self-regard, on the other hand, can manifest itself either as feelings of grandiose self-esteem or in the converse, as feelings of exaggerated inferiority.

In the present study, the self-regarding attitudes of a group of rebellious adolescents were compared to the self-regarding attitudes of an equated group of cooperative adolescents. Realistic self-regard was defined as moderate self-esteem. It was measured operationally by scores clustering around the mean of the distribution obtained on an instrument designed to measure self-esteem.

Narcissistic self-regard was defined as immoderately high self-esteem or extremely low self-esteem. It was measured by extreme scores on either the high or the low end of the distribution of scores obtained by means of an instrument designed to measure self-esteem.

The hypothesis of impaired ego-strength in the rebellious adolescent allowed the deduction that they will exhibit narcissistic attitudes toward themselves. Such attitudes were defined both as immoderately high self-esteem and as extreme feelings of inferiority. From this a second prediction follows: On an adjective list, checked for socially valued attitudes of the self, the rebel will tend to obtain extreme self-esteem scores, revealing feelings of exaggerated worth or inferiority.

The third hypothesis derived from the assumption of ego-weakness in the rebel is that of inadequate ego-ideal formation. Psychoanalysts regard the superego as an outgrowth of the ego, an outgrowth which depends upon the ego's tension-binding function, which serves to delay impulse gratification until such gratification is compatible with demands imposed by society (Fenichel, 1954). Ordinarily it is the parents who first mirror the society and who transmit societies' codes and expectations through the application of rewards and punishments. According to psychoanalytic theory, the superego results from the child's emotional acceptance and subsequent internalization of parental demands and prohibition. Such an acceptance and internalization requires that the ego be strong enough to renounce present pleasures for the sake of later (parental) approval. In the child this capacity to renounce is acquired as a consequence of the ego's developing ability to tolerate tension.

But the ego of the rebellious adolescent is weak and cannot tolerate tension. The rebel is impulsive. His impulsiveness demonstrates his ego weakness. If his ego is too weak to bind tensions it should also be too weak to allow the internalization of parental demands which constitute the growth of the superego (Zucker, 1943). Thus it can be expected that the superego formation of rebels is defective. That is the hypothesis tested here.

For the purpose of this experiment, only one aspect of superego defect was investigated. It was the assessment of the formation of the ego-ideal, i.e., the formation of the conscious part of the superego. Ego-ideal was defined as the complex of positive cultural values, which stem from the parents and represent them, and which have been accepted by the child. These serve as a model which the child strives consciously to emulate. An attitude of positive esteem for the parents and their perceived qualities is implied in this definition.

The status of a child's ego-ideal was measured in terms of the relationship between the subject's esteem for his parents and his own self-esteem. A strong ego-ideal formation was defined as a parent-esteem/self-esteem ratio equal to or greater than unity on an instrument designed to assess culturally valued attributes of the self and the parents. A weak ego-ideal formation was defined as a parent-esteem/self-esteem ratio smaller than unity on the same instrument. The hypothesis of defective ego-strength in the young rebel led to the prediction of the formation of a weak superego as measured by our instrument and defined as an ego-ideal score lower than unity. Prediction 3 stated that: On an ego-ideal measure, the rebel's self-esteem will tend to exceed his esteem for his parents, i.e., he will tend to obtain a parent-esteem/self-esteem ratio below unity.

The Sample

The *S*s used in this study were 75 high school students from the freshman and sophomore classes of a suburban high school. Their modal age was 14 years, with a range from 13 to 16 years. Their intelligence was average, with a mean at the 54th percentile on the American Council of Education test battery. The social status of the *S*s' parents was predominantly middle class.

Procedure for categorization. In order to divide the preliminary sample into the rebellious and cooperative groups, independent observations of the students were obtained. On the basis of these observations, the sample was divided into three groups according to the severity of the discipline problem which the *S*s presented to the high school authorities. For the final sample of 75 *S*s, only the two extreme groups were retained, the middle group was eliminated.

Three sources of information were utilized to establish criteria for placement of each *S* into one of the three groups: the rebellious group, the moderate group, and the cooperative group. One source was the *S*s' citizenship grade obtained during an 18-week interval. For each *S*, 18 ratings were available (three ratings by each of his six teachers).

The second source of information was the teachers' rankings of all the students in the freshman and sophomore classes on a continuum of rebelliousness. The third source of information was the dean's black list. On that list were found the names and the offenses of those students who had not responded to the teachers' disciplinary measures. This group of pupils represented the most serious disciplinary problems calling for special handling.

The three scores obtained from trichotomizing the distributions of citizenship scores, teachers' rankings, and presence on the black list were combined for each *S*. A new frequency distribution was made for the combined scores. *S*s whose combined score was in the highest third were classified as rebels. *S*s who obtained a zero score were designated as cooperators. The remaining middle group was eliminated from the final sample. The final sample of 75 *S*s comprised 30 cooperators and 45 rebels. No significant differences were found between the age and the intelligence of the *S*s in the rebel and cooperator groups. Nor were any significant differences found in the occupational status of their parents.

The Instruments

The Ego-Control Questionnaire

To measure ego-strength as a function of the individual's capacity to bind tensions, i.e., to control impulse gratification, a questionnaire was constructed. The method of its construction, the reliability and validity are discussed in full by Luft (1957). A few examples from the ego-control questionnaire are given below: "I sometimes open presents before I'm supposed to." "I usually make a plan before I start to do something." "I am always calm and cool." "I lose my temper easily."

The Adjective Check List

To assess self-regarding attitudes and ego-ideal formation, an adjective check list was devised. Each adjective included in the list had to meet the following criteria: (*a*) it had to be intelligible to the average freshman, (*b*) it had to describe personality unambiguously either favorably or unfavorably in the areas of intellectual, emotional, interpersonal, and physical functioning, (*c*) it had to differentiate between *S*s, and (*d*) it had to differentiate between the *S*s and their fathers and mothers.

Pretests on 72 high school sophomore and junior students yielded a preliminary pool of 151 adjectives. These were screened by consulting the Thorndike-Lorge word list to eliminate all those adjectives which were beyond the understanding of the average 13-year-old. In another pretest, the adjectives were further screened for intelligibility: the 72 pretest students were asked to define each of the randomized adjectives by making sentences out of them. Only those words which the students had been able to define correctly were retained in the revision list.

To establish the positive or negative value connotations of the adjectives empirically, 26 high school juniors were asked to judge each word. Eighty per cent agreement between *S*s as to favorable implication had been decided upon as the cutting point. Accordingly, each word for which there was less than 80% agreement was eliminated from the list as being too ambiguous a value statement to allow a clear differentiation of its positive or negative connotation.

The remaining 100 adjectives were pretested to weed out all those which did not differentiate sufficiently between *S*s. An item analysis of the responses to the adjectives was made. Twenty-five adjectives which produced the same response in all of the *S*s were rejected. Seventy-five items were retained in the final form of the adjective check list. The final adjective check list was administered to two freshman classes ($N = 63$) to pretest its discriminatory power for three test conditions: (*a*) self description, (*b*) description of fathers, and (*c*) description of mothers. It was found that the final form of the adjective check list differentiated adequately under all three conditions.

The following 75 words constituted the final check list:

Well Dressed
Confident
Thoughtful
Slow
Humorous
Insecure
Enthusiastic
Angry
Cooperative
Ashamed
Selfish
Quarrelsome
Cheerful
Helpful
Sulky
Creative
Considerate
Good Looking
Popular
Attractive
Dull
Highstrung
Respected
Wise
Complaining
Timid
Jumpy
Impatient
Narrow Interests
Graceful
Jolly
Blue
Sunny Disposition
Skillful
Easily Hurt
Nagging
Moody
Inventive
Tired
Awkward
Merry
Calm
Alert
Resentful
Athletic
Even Tempered
Original
Loving
Intelligent
Warm
Foolish
Worried
Irritable
Confused
Bossy
Nervous
Bright
Messy
Vigorous
Easy Going
Obeying
Generous
Strong
Relaxed
Sociable
Sympathetic
Catch on Quickly
Gentle
Well Groomed
Clear Thinking
Contented
Clumsy
Agreeable
Artistic
Good Natured

The adjective check list was administered to the experimental sample under three conditions:

Condition A: Each *S* was asked to mark as true the adjectives which described him as he really was.

Condition B: Each *S* was asked to mark as true all the adjectives which described his father as he (the S) saw him.

Condition C: The instructions for Condition C were to mark as true those items which described the S's mother correctly.

The adjective check list was scored by counting separately the number of positive attributes by each S under each of the three conditions. Three total esteem scores were thus available for each S. The self-esteem was measured by the score obtained under Condition A. The father-esteem was measured by the score obtained under Condition B, and the mother-esteem was measured by the score obtained under Condition C.

The Ego-Ideal Measure

For the purposes of this study, ego-ideal was defined as the ratio between the S's esteem for his parents and the S's self-esteem. Weak ego-ideal was thus defined as a parent-esteem score (P-E) lower than the self-esteem score (S-E); in other words, a weak ego-ideal was equated with a P-E/S-E ratio below unity. A strong ego-ideal was equated with a P-E/S-E ratio equal to or above unity.

In order to obtain the ratio between the Ss' parent-esteem and his self-esteem scores, the number of favorable adjectives checked for the father and the mother under Test Conditions B and C were added. This sum constituted the numerator in the P-E/S-E ratio. The denominator was the number of favorable adjectives the S had checked for himself under Test Condition A. A ratio falling below 2.00 denoted weak ego-ideal. A ratio equal or above 2.00 denoted strong ego-ideal.

Table 1

Chi Square Table of the Relationship Between Low Ego-Control and Rebelliousness

Ego-Control Scores	Cooperators	Rebels	Total
10–31 UC	10	29	39
32–46 OC	16	15	31
Total	26	44	70

Note.—$r_t = .43$; $\chi^2 = 4.99$; $.05 > p > .02$.

Results

Relationship between tension-binding-capacity and rebelliousness. This study's basic assumption of ego-weakness in the rebel led to the hypothesis of the rebel's decreased tension-binding-capacity. Operationally, tension-binding-capacity was defined as the ratio between impulse gratification vs. delay of gratification (IG/DG ratio). It was predicted that on a questionnaire designed to measure the IG/DG ratio the rebel will be on the under-controlling end of the distribution, i.e., he will obtain lower scores.

The ego-control questionnaire was administered to 26 cooperators and 44 rebels. The tetrachoric correlation was computed as a measure of the degree of association between rebelliousness and under-control. Statistical analysis yielded an $r_t = .43$. The chi square was 4.99, which for one degree of freedom is significant between the .05 and .02 levels of confidence.

This degree of statistical significance supports the hypothesis that decreased tension-binding-capacity as defined here and rebelliousness are correlated.

Relationship between "unrealistic" self-esteem and rebelliousness. The hypothesis of decreased tension-binding-capacity in the rebel permitted the inference that the rebel will show deficiencies in reality-testing as, for instance, in his capacity to evaluate himself realistically. Prediction 2 stated that the rebel's unrealistic self-regard would be revealed on the adjective check list in extreme self-esteem scores, indicating feelings of exaggerated worth or inferiority.

The obverse was predicted for the cooperator: his more realistic self-regard was expected to yield moderate self-esteem scores.

The adjective check list was administered to 30 cooperators and to 45 rebels. It was scored by counting for each S the number of positive adjectives checked. The difference in variance of the two distributions of scores was tested. The obtained sigma for the cooperators was 8.51; the sigma for the rebels was 12.71. The observed difference was in the predicted direction and yielded an F ratio of

2.21. For 44 *df* and 29 *df*, respectively, the obtained *F* ratio is significant beyond the .01 level of confidence. This degree of statistical significance supports the hypothesis that "unrealistic" self-esteem as defined by extreme scores on a self-esteem measure is a function of rebelliousness.

Relationship between rebelliousness and ego-ideal weakness. A weakened ego-ideal in the rebel was predicted from the hypothesis of impaired ego-strength underlying asocial behavior and the consequent defective superego functioning. Prediction 3 stated that the ego-ideal ratios of rebellious adolescents would tend to be less than 2.00, or unity, more frequently than the corresponding ego-ideal scores of cooperative adolescents, which would be found more frequently above unity, i.e., be equal to or greater than 2.00.

The adjective check list was administered to 25 cooperators and to 39 rebels.[3] By summing the father- and mother-esteem scores and dividing the result by the self-esteem score ($F + M/S$), the "ego-ideal" ratio was computed for each *S*.

In order to show that the rebels, more frequently than cooperators, tend to reject their parents as ego-ideal, chi square was computed. The obtained chi square of 3.05 was significant beyond the .05 level of confidence (one-tailed test), for 1 *df*. The hypothesis that rebels have a weaker ego-ideal is supported by the data. The results are summarized in Table 2.

Table 2

Chi Square Table of the Relationship Between Weak Ego-Ideal and Rebelliousness

Ego-Ideal Ratio	Cooperators	Rebels	Total
1.00–1.99	5	16	21
2.00–4.50	20	23	43
Total	25	39	64

Note.—$\chi^2 = 3.05$ (predicted direction); $p < .05$.

Discussion

A great deal of knowledge has accumulated regarding the determinants of asocial behavior in the juvenile delinquent. Much less is known about the "average" rebellious adolescent who has perpetrated similar misdemeanors but has not come to the attention of the law. The study presented here was concerned with the rebel before, rather than after, social intervention has branded him as a member of a special outgroup. Our rebellious adolescents were at the time of the study in regular attendance at school. Although they were known to their teachers and counsellors as difficult to handle, they were more nearly representative of the adolescent population than the "legal delinquents" upon whose behavior much of the research on asocial behavior has been based. The study of the rebel was conceived of as contributing a link in the continuum of social behavior, ranging from the extremely disruptive to constructive activities. The psychoanalytic construct "impaired ego-strength" presented itself as a convenient unifying principle, which allowed us to order divergent behavior manifestations into one system. The positive findings of this study substantiated psychoanalytical theory at several levels of abstraction. By extrapolating from the data, the primary explanatory principle was confirmed to the extent that the study affirmed the lower order hypotheses derived from the construct "impaired ego-strength."

There is no doubt that alternate hypotheses of the observed results can be proposed. These are certain to occur to the critical reader and are therefore not offered here. Many possibilities for extending the present research will also come to mind. Of these only one is mentioned here: the argument that the rebels were correct in their negative evaluation of their parents cannot be met directly in the absence of any objective descriptions of the two sets of parents. It may be that dispassionate observers would tend to agree with the rebels that their parents were less admirable than parents of cooperators; but perhaps not. In the event that the rebels were correct in stating that their parents had few socially desirable traits, one could then ascribe their defective superego to the inadequacies of their models and the conflict between their parental models and the ideals held up by the society in which the rebels live. If that is the case, superego development must be viewed in terms of the actual failing of parents to provide worthy behavior models to the child. On the other hand, should it be shown that objective evaluation of the rebels' parents proves them not to differ from other parents in the possession of socially admirable traits, then the qualities attributed to parents by rebellious offspring represent their own peculiarly dour views. Whether such views would originate from hostility, disappointment, or projection remains to be investigated.

---

[3] The number of *S*s decreased from one set of conditions to another. Not all of those who had filled out the adjective check list under Condition A were willing to continue under Conditions B and C. The attrition rate was six *S*s, one cooperator and five rebels.

The evaluation of the parents of rebels beckons as the next research step.

Summary and Conclusions

A sample of high school students was divided into 45 rebels and 30 cooperators. A questionnaire was constructed to measure ego-strength defined as tension-binding-capacity. Rebelliousness was correlated with decreased tension-binding-capacity. An adjective check list was developed. On it rebels revealed unrealistic self-regard, which differed from the realistic self-appraisal of the cooperators. Rebels were found to regard themselves more highly than they regarded their parents; cooperators admired their parents more than themselves.

The results supported the psychoanalytic thesis of ego-weakness in the rebel. Implications for further research were discussed.

Received January 21, 1958.

REFERENCES

Aichhorn, A. *Wayward youth.* New York: Viking, 1936.

Block, J. An experimental investigation of the construct of ego-control. Unpublished doctoral dissertation, Stanford Univer., 1950. Pp. 113 ff.

Fenichel, O. *The psychoanalytic theory of neurosis.* New York: W. W. Norton, 1954.

Johnson, Adelaide M. Sanctions for superego lacunae of adolescents. In K. R. Eissler (Ed.), *Searchlights on delinquency. New psychoanalytic studies.* New York: Int. Univer. Press, 1949. Pp. 225–245.

Lampl-de-Groot, Jeanne. Neurotics, delinquents and ideal-formation. In K. R. Eissler (Ed.), *Searchlights on delinquency. New psychoanalytic studies.* New York: Int. Univer. Press, 1949. Pp. 246–262.

Luft, Joseph. Psychological control patterns within families. *J. consult. Psychol.,* 1957, **21**, 206.

Redl, F., & Wineman, D. *Controls from within: Techniques for the treatment of the aggressive child.* Glencoe, Ill.: Free Press, 1952.

Shippee, Eva-Maria. Ego-structure of rebellious adolescents: An experimental study of the dynamics and etiology of asocial behavior. *Dissertation Abstr.,* 1954, **14**, 998–999.

Zucker, H. J. Affectional identification and delinquency. *Arch. Psychol.,* 1943, **25**, 1–50.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

AN APPROACH TO THE FACTOR STRUCTURE OF CLINICAL JUDGMENTS

MARTIN D. CAPELL AND JULIAN WOHL
VA Mental Hygiene Clinic, Detroit

There is a growing awareness in the clinical field that the process of describing, rating, or classifying a person is extremely complex and that there are many determinants of the clinical impression or judgment. Previous studies (Buss & Gerjuoy, 1957; Grayson & Tolman, 1950) of concepts used by clinicians have not been concerned with either the clinical judgment itself or with attempting empirically to dimensionalize the concepts. Psychologists tend to be critical of their concepts and suspicious of themselves inasmuch as they have come to recognize that, in making judgments about other people, they do so not only as scientists but also as human beings. Recent interest in the utilization of countertransference in the psychotherapeutic process (Cohen, 1954) is one aspect of this concern. Osgood (1953), in his work with the semantic differential, found that the judgments of words by groups of college students could be accounted for on the basis of three factors: evaluation, activity, and potency. The evaluative ("good-bad") factor alone accounted for approximately 50% of the total variance in the judgments. In the present study, the factor structure of clinical judgment is investigated with particular interest being given to the role of the evaluative factor in such judgments.

Method and Procedure

The psychiatric, social work, and psychological staffs at the Detroit VA Mental Hygiene Clinic, a total of 28 persons, were approached individually and requested to give 10 of the concepts they commonly used in describing patients. Following, each person was asked to indicate the "opposite" of each of the concepts he or she had contributed.

The obtained concept-dimensions were categorized by the *E*s. Twelve dimensions were selected on the basis of frequency of occurrence (four or more). These were as follows:

anxious–calm
narcissistic, self-invested–object-invested
inhibited, rigid, reserved–expansive, expressive, outgoing
orally demanding–realistic help-seeking
flat, apathetic, affectless–sensitive, emotionally responsive
defensive, guarded–open, free of defensiveness
distorted reality testing–undistorted reality testing
precarious, transient interpersonal relationships–stable, firm interpersonal relationships
hostile, unaccepting, suspicious–friendly, trusting, accepting
passive–active, aggressive
dependent–independent
inadequate–adequate, capable

Three of the dimensions which Osgood found to load highly on the evaluation factor were added:

pleasant–unpleasant
honest–dishonest
kind–cruel

Finally, five typical psychoanalytic constructs:

much secondary gain–little secondary gain
positive Oedipal strivings–negative Oedipal strivings
genital character–pregenital character
many anal character traits–few anal character traits

Table 1

Rotated Factor Matrix Derived from Scale Intercorrelations

Scales[a]	Factor[b] I	II	III	IV
Anxious	09	61	58	07
Narcissistic, self-invested	−36	67	−02	07
Inhibited, rigid, reserved	−57	45	−09	54
Pleasant	85	05	−40	−25
Orally demanding	−72	00	06	30
Flat, apathetic, affectless	−51	46	12	46
Defensive, guarded	−43	45	57	06
Much secondary gain	−26	24	−26	−08
Positive Oedipal strivings	32	−35	−02	15
Distorted reality testing	−80	21	28	−18
Precarious, transient interpersonal relationships	03	74	51	−32
Kind	65	−03	00	33
Genital character	40	−26	24	03
Hostile, unaccepting, suspicious	−17	82	00	54
Many anal character traits	03	33	36	04
Passive	−34	61	15	48
Many oral character traits	−06	61	48	17
Dependent	−08	94	05	01
Honest	50	−01	45	06
Inadequate	−12	68	−12	09

Note.—Decimal points have been omitted in the factor loadings.

[a] Only the positive ends of the scales are reproduced below. Consult the text for the complete scales.

[b] The mean squared factor loadings are as follows: I, .198; II, .258; III, .096; and IV, .075.

many oral character traits–few oral character traits

were included. The 20 dimensions were then ordered randomly and reproduced on a duplicated form as a series of seven-point scales. Following a presentation of the case of a patient to one of the clinic's consulting psychoanalysts, every person in attendance (excepting the *E*s) was asked to rate the patient on the scales. The professional discipline of the judge was also obtained, but the ratings were otherwise anonymous. The judges consisted of 3 psychiatrists, 11 clinical social workers, and 2 clinical psychology trainees, a total of 16 persons.

Product-moment intercorrelations among the scales were obtained, and a centroid factor analysis performed on the resulting matrix. Four factors were extracted. In order to determine the extent to which values did enter into the ratings, the first factor was rotated through the locus of the points which defined the three evaluative scales. The remaining factors were rotated blindly and orthogonally.

Results and Discussion

The rotated factor matrix is given in Table 1. The scale loadings on Factor I, the factor in which the evaluative component in the ratings was maximized, suggest that a number of clinical concepts have an evaluative meaning along with their diagnostic intent. In particular, such concepts as *distorted reality testing, orally demanding, inhibited,* and *flat* appear to reflect in large part "bad" characteristics, while *genital character* and *positive Oedipal strivings* have "good" implications. In the light of this evidence, caution in the use of these terms in clinical parlance might well be considered.

The mean squared factor loading on Factor I (.198) constitutes approximately 30% of the common factor variance obtained. This can be contrasted with the results obtained by Osgood (1953), where roughly 70% of the common factor variance was evaluative. Thus, as compared with college students performing semantic ratings, clinical judgment appears to be relatively unbiased. This comparison needs to be limited by the fact that the ratings in the present study are presumably denotative, rather than connotative as in Osgood's study.

Factor II, orthogonal to the evaluative factor, appears to be a dimension of psychopathology and contains about 47% of the common factor variance. It suggests that clinical judgment of pathology is predominantly unbiased by evaluative considerations. Factors III and IV do not appear to be as readily interpretable. Factor III is tentatively identified as a dimension referring to the person's desire and need for psychotherapeutic intervention. Factor IV appears to be an inhibition or social withdrawal dimension.

The present study can be considered only as exploratory, considering the small *N* involved. Nevertheless, the obtained factors follow along lines that might be expected in an outpatient, psychoanalytically oriented, psychotherapeutic agency. The clinic personnel look at the range and degree of psychopathology, using generally psychodynamically derived concepts rather than those coming from descriptive psychiatry. They judge people as to their need and wish for treatment and with regard to their accessibility to a psychotherapeutic approach. A number of their concepts are endowed with considerable evaluative nuance.

Summary

The factor structure of clinical judgments was explored with particular interest being placed upon the role of values in the making of such judgments. A group of mental hygiene clinic personnel was asked to rate a patient on a number of scales derived from commonly used clinical concepts. Three "evaluative" scales and five psychoanalytically derived scales were also included. The scales were intercorrelated and a factor analysis performed. The first factor was rotated through the evaluative scales. Less than one third of the common factor variance was accounted for by this factor; it suggests that evaluation does play a part in clinical judgment. The remaining three factors were tentatively identified as dimensions of psychopathology, need and wish for psychotherapy, and social inhibition.

Received January 22, 1958.

REFERENCES

BUSS, A. H., & GERJUOY, G. The scaling of terms used to describe personality. *J. consult. Psychol.*, 1957, **21**, 361–369.

COHEN, MABEL B. Countertransference and anxiety. *Psychiatry*, 1954, **17**, 231–243.

GRAYSON, H. M., & TOLMAN, RUTH S. A semantic study of concepts of clinical psychologists and psychiatrists. *J. abnorm. soc. Psychol.*, 1950, **45**, 216–231.

OSGOOD, C. E. Report on development and application of the semantic differential. Unpublished manuscript, Univer. of Illinois Library, 1953.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

BODY REACTIVITY GRADIENTS AND FIGURE DRAWING VARIABLES[1]

SEYMOUR FISHER
Baylor University College of Medicine

There is now considerable literature (Fisher & Cleveland, 1955; Fisher & Cleveland, 1958; Witkin, Lewis, Harztman, Machover, Meissner, & Wapner, 1954) devoted to the study of body image problems. It has become apparent that the individual's attitudes toward his body not only supply valuable information about his past socialization experiences but also provide a basis for predicting important aspects of his behavior. Body image measures have been successfully used to clarify such diverse phenomena as the appearance of phantom limb sensations following amputation (Haber, 1956); the content of schizophrenic delusions and hallucinations (Fenichel, 1945); differential capacity to deal with stress (Fisher & Cleveland, 1958); and the ability to make spatial judgments in unstructured settings (Witkin et al., 1954). In earlier work the writer focused particularly on the possibility of predicting individual patterns of physiological reactivity in terms of body image variables. This work aimed to verify the general proposition that the individual's attitudes toward his body may influence its physiological reactivity. Data were obtained which indicated that certain patterns of physiological response can be predicted from body image variables. Thus, it was established (Fisher & Cleveland, 1957; Fisher & Cleveland, 1958) that the more definite an individual's body image boundaries the more likely he is to show relatively greater reactivity in the outer body layers than in the body interior. More recently, observations were made which indicate that individuals manifest differences in the GSR reactivity of their right and left body sides and that the pattern of these differences is linked with body image factors.

It is with this problem of right vs. left reactivity that the present paper concerns itself. Previous findings concerning this phenomenon have been as follows (Fisher, 1958; Fisher & Abercrombie, in press):

1. Right-handed subjects (*S*s) who, while wearing aniseikonic lenses, clearly distinguish the left body side as smaller than, or inferior to, the right side in their body image evaluations are significantly more likely than those not making this distinction to manifest a GSR gradient such that the left side is more reactive than the right side.

2. Right-handed *S*s who are secure about their over-all body images (as measured by reaction to tachistoscopically presented picture of mutilated bodies) show relatively greater left than right GSR reactivity; whereas *S*s who are insecure about their body images either show no gradient at all or relatively more right than left reactivity. The optimum GSR gradient associated with a well integrated body image is the left directional one.

3. Left-handed *S*s fail to evidence significant relationships between body image measures and the GSR gradient variable.

A primary goal of the present project was to determine how these findings would stand up when approached in terms of new body image measures derived from figure drawings. Figure drawings were chosen for special study in this respect because the Draw-A-Person method has in the past been used more than any other measure for evaluating

---

[1] This study was supported by a Public Health Career Investigator grant (M-1578).

body image variables (Abel, 1953; Bender & Keeler, 1952; Fisher & Cleveland, 1958; Silverstein & Robinson, 1956; Witkin et al., 1954).

The attainment of the first primary objective resolved itself into two specific problems:

1. To determine if GSR gradient patterns are linked with the right side vs. left side attributes of figure drawings. The specific prediction would be that those *S*s who depict the left side of the figures as smaller than the right side would show relatively greater left GSR reactivity than right reactivity. Those drawing the right side of the body as relatively smaller or who depict no differences between the two body sides would fail to show the optimum left directional GSR gradient.

2. To determine if GSR gradient patterns are related to indices of over-all body image integration derived from figure drawings. Here the specific prediction would be that *S*s whose drawings evidenced good body image integration would be more likely to be typified by the norm left reactive GSR gradient than *S*s whose drawings evidenced poor body image integration.

A second primary intent of the study was to test, by means of figure drawings, an hypothesis proposing that the existence of an optimum GSR gradient between the two body sides is at one level a function of whether the individual has learned a clear distinction between male and female sex roles such that the male is considered superior in size and strength. This hypothesis was the product of a rather speculative train of logic. It grew out of a proposition, which has been much elaborated elsewhere (Fisher & Cleveland, 1958), that an individual's body image attitudes do not reflect the actual appearance of his body but rather express or summarize patterns of important socialization experiences. This point was well exemplified in the finding that the definiteness which an individual ascribes to his body boundaries is not correlated with his actual physique as measured by body type but is significantly linked with the definiteness of certain roles and values he learned from the parental figures (Fisher & Cleveland, 1958). Findings of this sort suggested that body image attitudes were primarily a projection onto a figurative "body screen" of important roles and concepts which had been learned.

When, therefore, differences were found in the size and strength attributed by individuals to their two body sides, and when these differences proved to be correlated with GSR gradients of reactivity, the question arose as to what underlying socialization variable or experience was being thus reflected. There were already observations in the literature (Fenichel, 1945; Wolff, 1943) which implied that attitudes toward the two body sides were influenced by conflicts about masculinity-femininity. It had been noted that in some dramatic instances schizophrenic individuals would act out sex role conflicts by developing the delusion that one side of their bodies was male and the other side female. Sufficient leads of this kind could be found so that it seemed logical to relate differences in right-left body image and the correlated GSR reactivity differences to sex role variables. One could reason that the definiteness of an individual's right-left body image and GSR differences was correlated with how well he had learned a stable nonconflictual distinction between the masculine and feminine roles. However, it seemed doubtful that a distinction which attributed relatively greater power to the female than to the male and which therefore was opposed to the usual cultural stereotype of masculine-feminine size relationships could be maintained by the individual without a great deal of anxious confusion. So the final hypothesis which emerged proposed that a definite and optimum GSR reactivity gradient between the two body sides was based on a significant body image distinction in the size attributed to the two body sides and that this distinction was in turn dependent upon the individual having learned to discriminate clearly between the male and female sex roles in such a fashion that the male was considered to have greater size and power.

Procedure

Measure of Right-Left Body Reactivity

The method for measuring relative right-left GSR reactivity has already been described in detail elsewhere (Fisher, 1958; Fisher & Abercrombie, in press) and therefore will be only briefly outlined. It is of the exosomatic type and involves simultaneous recordings from the left and right sides on separate channels against a relatively inactive upper arm. Resistance changes across the skin are picked up as changing potential and fed into a Grass EEG power amplifier. Recordings were taken simultaneously: first from the middle fingers; secondly from the forefingers; then from the fingers next to the little fingers; and finally from two corresponding points on the palms. The currents to each finger were carefully matched. Also, a common calibrating resistor in series with the upper arm checked the equivalence of the sensitivities of the two circuits. The area of recording from each site was controlled by masking off all but a fixed circular region by means of punched tape. For each *S* there was finally available side by side the continuous GSR responses for each set of right-left sites. The amplitudes of response at two homologous sites at a given time could then be measured and compared. If the differences occurring at a given site were such that twice as many favored one side over the other and if at least eight were present, then it was judged that a directionality for that site has been demonstrated. When an *S* showed this kind of directionality in favor of one side for three out of four sites, he was considered to have a definite GSR gradient in that direction.

Figure Drawing Measures

Each *S* was asked to draw a full-length picture of a person. When he had finished his first sketch, he was asked to draw a person of the opposite sex. The usual instructions were modified to the extent that *S*s were told to draw only front views of each figure. It was necessary to do this in order to have both the right and left sides of each figure fully represented.

The following measures were derived from the drawings:

A. *Right-Left Size Differences.* Size differences between the right and left sides of each drawing were determined on the basis of the relative dimensions of the right and left arms and also of the right and left legs. Arm and leg differences were chosen not only because they are often the most prominent indicators of asymmetry in figure drawings but also because other indices involving the trunk and head were highly unreliable due to the difficulty in establishing "middle" reference points from which to make right vs. left measures. The length of each of the figure's arms was measured with a protractor from a point half way between the indicated boundaries of the uppermost part of the arm to a point at the tip of the longest finger. The length of each leg was measured from a point half way between the boundaries of the leg at the point where it joined the middle of the body (crotch area) to a point indicated by the most distant toe. This leg measure was made only on the drawing of the man because in the instance of the female drawing the dress almost always concealed the legs.

B. *Over-all Body Image Disturbance.* This measure was derived from one which was previously used by Witkin et al. (1954). It was developed by Machover in an attempt to predict on the basis of the individual's body concept how well he would perform on tasks requiring spatial judgments in very unstructured situations. Machover compiled an extensive list of about 40 figure drawing signs which she considered on the basis of clinical experience to indicate lack of body confidence and general difficulty in developing an acceptable body concept. Scores computed from these signs proved to differentiate significantly between *S*s who presumably could use their bodies adequately as a frame of reference for making spatial judgments and those who could not do so. In reviewing Machover's list of signs, it became apparent that many of them involved making rather complex subjective judgments. It was therefore decided to use only 14 of the signs which were based on simple and relatively objective judgments. One penalty point, indicating disturbance in body concept, was assigned for the presence of each of the following signs:

1. Erasures.
2. Transparency such that the figure defies the laws of perspective as regards the masking of objects when they are behind others.

3. Lack of any body part.
4. Nose indicated only by two nostril dots.
5. Mouth indicated only by a line.
6. One or more arms behind back.
7. Very crude or peculiar clothing.
8. Lack of breasts in the female figure.
9. Shading of the body.
10. Lack of margins and delimiting lines in the figure (e.g., cuffs, collar, hemline).
11. Figure markedly off balance.
12. Figure very small (less than one-half the length of the page).
13. Markedly unusual shading or elaboration of the crotch area.
14. Opposite sex drawn first.

A total score was tabulated which equaled the sum of the penalty points assigned to both of the figures. It will be referred to as a "Body Disturbance" score. The scoring was done blindly with all subject identification removed.

Subjects

The *S*s consisted of 34 men and 16 women. Their median age was 21 and their median educational level was 14 years. They were mainly students who were recruited from college organizations by paying each organization a fee to have its total membership participate in the experiment. Only right-handed *S*s were used because previous work (Fisher, 1958) had indicated that due to complicating factors the relationships of body image and GSR variables are masked in left-handed individuals.

C. *Relative Size Attributed to Masculinity vs. Femininity.* The assumption was made that the individual's concept of the relative size and power attributes of the male vs. female roles would be indicated by the relative heights of the male and female figures he drew. This was purely an operational procedure. Each drawing was evaluated by measuring from the tip of the head to the furthest distant point on one of the legs. Only when the male figure was at least one-half of an inch larger than the female figure was the difference considered significant. This criterion was adopted in order to minimize chance differences in size which might result from stray pencil stroke extensions.

Results

Tabulation of the data indicated that 21 *S*s were left directional in their GSR gradients; 9 were right directional; and 20 manifested no gradient at all. A chi-square analysis demonstrated only a chance relationship between directionality of GSR and figure drawing right-left side size differences. This held true for the figure drawing arm measures considered separately and also for the leg measures. Figure drawing asymmetry seems to bear no relationship to GSR response asymmetry.

However, a significant relationship in the predicted direction was found by means of chi square between the figure drawing Body Disturbance scores and GSR directionality. *S*s scoring below the median in Body Disturbance significantly (.001 level) more often manifested left directional GSR gradients than *S*s above the median in Body Disturbance. Those above the median more often showed either no GSR gradient at all or right directionality of response.

Eighteen *S*s drew the male figure as larger than the female figure. Thirty-two drew the female figure as equal in size or smaller than the male figure. There were no sex differences among the *S*s in this respect. Chi-square analysis of the relationship of the GSR gradient variable to the sex size difference variable indicated a significant relationship in the predicted direction. Those *S*s who manifested the left reactive GSR gradient drew the male figure as larger than the female figure significantly (.05 — .02 level) more often than did *S*s who had no GSR gradient or who were right reactive.

Discussion of Results

A significant link between GSR directionality and asymmetry of the figure drawings was not demonstrated. This suggests that figure drawings do not tap body image asymmetry attitudes in the same fashion as the aniseikonic lens technique (Fisher, 1958), which was first used to demonstrate a significant relationship between body image asymmetry and GSR right-left gradients. The aniseikonic lens measure required *S*s to make a series of size comparisons of their homologous right and left fingers. This was done under conditions in which their perceptual fields were so distorted as to encourage projection of autistic attitudes. The aniseikonic lens method for evaluating body image asymmetry contrasts with the figure drawing method in two ways. First of all, it requires direct rather than indirect judgments about the individual's own body. Secondly, it sets the specific task for the individual of making judgments regarding symmetry; whereas the figure drawing asymmetry measure is derived from responses to a task having no specific symmetry connotations. This implies that if one wants to measure body image attitudes toward very specific areas or dimensions of the body, such a measure is most likely to be successful if it requires direct evaluation of these areas of the *S*'s own body.

However, the results obtained relative to the figure drawing Body Disturbance score indicate that such specific evaluation by the individual of his own body is not necessary in order to measure more general and abstract body image dimensions. That is, a significant relationship was found between the Body Disturbance score and the GSR gradient variable. Similarly, in a previous study (Fisher & Abercrombie, in press) it was possible to establish a significant relationship in the same direction between GSR gradient and an index of general body image disturbance based on reactions to tachistoscopically presented mutilated figures. This index, like the figure drawing score, was an indirect measure which in no way required the *S* to make judgments about his own body.

The demonstration of a link between the Body Disturbance score and the GSR response pattern adds further validity to the proposition that body image attitudes influence the body's reactivity characteristics. But the most intriguing finding of the present study is that supporting the hypothesis concerning the relationship of GSR response patterning to definiteness of sex role distinction. Such a finding gives a bit of substance to the three-level general theory which has been proposed concerning role, body image, and physiological reactivity. There is now at least one empirical result, as modest and tentative as it may be, which indicates that a body reactivity pattern which has been shown to be related at one level to body image attitudes is at still another level related to the manner in which an individual conceptualizes an important life role. The presumption is that the body image attitude represents a translation into body terms of the role concept. But this remains to be proven. A study is now under way which attempts such proof by tracing the relationships of GSR patterns, body image attitudes, and role attitudes over a wide age range from early childhood through adulthood.

Summary

Previous work demonstrated relationships between body image measures and GSR right-left reactivity gradients. In the present study, body image measures derived from figure drawings failed to confirm a previous finding regarding the relationship of body image right-left size asymmetry and GSR directionality. However, a figure drawing measure of over-all body image disturbance was found to be linked in the predicted direction with GSR right-left responsivity. Further, a significant relationship was established between a figure drawing measure of the individual's concept of the relative size and strength attributes of the male and female roles and GSR reactivity gradients. The findings lend support to a general theory that relates role, body image, and physiological reactivity.

Received January 28, 1958.

REFERENCES

Abel, T. A. Figure drawing and facial disfigurement. *Amer. J. Orthopsychiat.*, 1953, **23**, 253–264.

Bender, L., & Keeler, W. R. The body image of schizophrenic children following electro-shock treatment. *Amer. J. Orthopsychiat.*, 1952, **22**, 335–355.

Fenichel, O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.

Fisher, S. Body image and asymmetry of body reactivity. *J. abnorm. soc. Psychol.*, 1958, **57**, 292–298.

Fisher, S., & Abercrombie, J. The relationship of body image distortions to body reactivity gradients. *J. Pers.*, in press.

Fisher, S., & Cleveland, S. E. The role of body image in psychosomatic symptom choice. *Psychol. Monogr.,* 1955, **69**, No. 17 (Whole No. 402).

Fisher, S., & Cleveland, S. E. An approach to physiological reactivity in terms of a body-image schema. *Psychol. Rev.,* 1957, **64**, 26–37.

Fisher, S., & Cleveland, S. E. *Body image and personality.* Princeton, N. J.: Van Nostrand, 1958.

Haber, W. B. Observations on phantom limb phenomena. *A.M.A. Arch. Neurol. Psychiat.,* 1956, **75**, 624–636.

Silverstein, A. B., & Robinson, H. A. The representation of orthopedic disability in children's figure drawings. *J. consult. Psychol.,* 1956, **20**, 333–341.

Thass-Thienemann, T. Left-handed writing. *Psychoanal. Rev.,* 1955, **42**, 239–261.

Witkin, H. A., Lewis, H. B., Hertzman, M., Machover, K., Meissner, P. B., & Wapner, S. *Personality through perception.* New York: Harper, 1954.

Wolff, W. *The expression of personality.* New York: Harper, 1943.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

POPULATION DIFFERENCES IN CONSTRUCT VALIDITY

GERALD S. LESSER
Hunter College

In the seldom definitive efforts to validate measures of personality variables, it frequently happens that a test of "construct validity" (American Psychological Ass., 1954) performed on a particular population will not be confirmed by an otherwise identical validating test performed on a different population. This phenomenon has appeared in a variety of research areas, including attempts to establish the construct validity of projective measures predicting response to stress (Carlson & Lazarus, 1953; Williams, 1947), projective signs of organic brain damage (Dorken & Kral, 1952; Fisher, Gonda, & Little, 1955), and measures of authoritarianism and anxiety (Davids, 1955; French, 1955). In each pair of studies, identical measures of personality variables were used, but different populations were employed and different results were obtained.

Interpretations of failures to demonstrate consistent validity with diverse populations ordinarily focus upon the limitations of the measure being validated. However, there have been few attempts to assign psychological meaning to the differences between the populations which produce the differential validity results. Why do certain populations yield positive validating findings, while other populations, in response to identical measuring instruments, yield inconclusive or negative results?

The answer proposed here is a special application to validation studies of the well-accepted general proposition that the relationships among personality measures will depend upon the context of variables in which the personality measures are obtained. The present study demonstrates that (*a*) a single measure may display differential validity for different populations, and (*b*) psychological variables may be identified which will allow prediction of the nature of the differential validation results for different populations.

The variable of anxiety was selected as one of the possible dimensions of population difference affecting validation findings. Empirical evidence (e.g., Grinker & Spiegel, 1945; Luria, 1942; Mahl, 1949) from diverse areas of psychological research links anxiety with disruption, disorganization, and erratic quality of behavior. Kuhlen emphasizes this relationship between anxiety and response variability, concluding that "anxiety will show in . . . shifts of mood, instability, unpredictability and . . . inconsistency in performance" (Kuhlen, 1952, p. 268). In addition, many clinical discussions have centered upon the influence of anxiety upon the validity of diagnostic test results.

The present study does not, however, deal directly with the effects of general, pervasive anxiety but rather with anxiety specifically related to the expression of aggression. The assumption is made that this more circumscribed anxiety will manifest certain consequences which resemble the effects of general anxiety, i.e., anxiety about aggression will be associated with inconsistency in expressions of aggression at different times and in response to different measurement techniques. Individuals and populations exhibiting high anxiety over the expression of aggression should manifest greater variability in aggressive response and, hence, should exhibit lower intercorrelations among measures of aggression. The construct validity of measures of aggression should be difficult to establish in such populations.

Direct assessment of anxiety about aggression, independent of aggressive need strength, encounters formidable methodological difficulties. Indirect methods are suggested by the literature. For example, convincing evidence (Dollard, Doob, Miller, Mowrer, & Sears, 1939; Hollenberg & Sperry, 1951; Miller, 1948; Whiting & Child, 1953) exists for the relationship between inhibitory socialization of aggression and the learning of anxiety about aggression. An indirect index of the strength of aggression anxiety in preadolescent boys was obtained by measuring the attitudes and practices toward aggression socialization of their mothers.

Hypothesis

The population difference examined in this study represents a difference in the conditions of learning to express aggressive responses and the resulting difference in anxiety about aggression.

The hypothesis tested is that, under conditions of high anxiety over the expression of aggression, the intercorrelations among various measures of aggression are significantly lower than under conditions of low anxiety over the expression of aggression.

Method

Subjects

The *S*s were 44 white boys (ages 10–0 to 13–2) and their mothers. The boys were drawn from one fifth grade and two sixth grades in two public schools. All of the boys and their mothers in these three classes participated, except one mother who refused to be interviewed. The Kuhlmann-Anderson intelligence quotients of the boys ranged from 82 to 119, with a mean of 102. The two schools are in adjacent districts and the families constitute a relatively homogeneous upper lower-class group.

Maternal Attitudes and Practices

Only one aspect of the environmental conditions of learning of aggressive behavior was measured, i.e., the maternal attitudes and practices supporting or prohibiting aggression. A structured questionnaire-interview schedule was orally administered to the mothers in their homes by a male interviewer. Questions regarding the support or prohibition of aggression constituted only one segment of the total interview; the entire interview schedule is described in detail elsewhere (Lesser, 1952). Pertinent to the present study were 8 items concerning the mother's attitudes toward aggression in children, and 13 items about the mother's practices in dealing with the aggressive behavior of her child. An illustrative item measuring maternal attitudes toward aggression is: "A child should be taught to stand up and fight for his rights in his contacts with other children." The four response alternatives of agree, mildly agree, mildly disagree, and disagree were allowed for this item. An example of an item measuring maternal practices concerning aggression is: "If your son comes to tell you that he is being picked on by a bully at the playground who is his own age and size, there would be a number of different things you might tell him. Would you tell him to ignore him and turn the other cheek?" Response alternatives for this item were Yes and No. Items that did not involve judgments on a four-point scale were transformed to have approximately the same range of scores as the items that involved four alternatives.

A single score was obtained for each mother by combining all items, assigning plus scores to the responses indicating support of aggression and minus scores to responses indicating discouragement of aggression. The range of scores was from $+9$ to -7, with a median score of $+2$. The corrected odd-even reliability coefficient was .80.

The distribution of scores for maternal response to aggression was dichotomized to form one group of mothers (with scores above or at the median) whose attitudes and practices were more supportive of aggressive behavior than those of the other group (with scores below the median). The hypothesis demands that the intercorrelations among various measures of aggression for the children ($N = 23$) of the mothers who support aggression be significantly more positive than the intercorrelations for the children ($N = 21$) of the mothers who oppose aggression. Boys whose mothers encourage aggression will be referred to as the Low Anxiety (LA) group; boys whose mothers discourage aggression will be called the High Anxiety (HA) group.

Measures of Aggression

Children's Form of Rosenzweig Picture-Frustration Study. Measures of extrapunitiveness (E), intropunitiveness (I), and impunitiveness (M) were obtained through standard group administration and scoring of the Children's Form of the Picture-Frustration Study (Rosenzweig, Fleming, & Rosenzweig, 1948).

The present sample of 44 boys differs significantly ($p < .05$) from Rosenzweig's normative group of 10- to 13-year-old boys, manifesting higher E scores (55% vs. 42%), lower I scores (21% vs. 28%), and lower M scores (24% vs. 30%).

The corrected matched-half reliability coefficients for the group of boys whose mothers encourage aggression (Low Anxiety group) were: E, .85; I, .58; M, .82. For the group of boys whose mothers discourage aggression (High Anxiety group), the corresponding reliability coefficients were: E, .95; I, .73; M, .70. Differences in reliability coefficients for the two groups were not statistically significant. These reliability coefficients are somewhat greater than those typically reported for the Picture-Frustration measures. This may be a function of the fact that the matched-half technique for computing reliability was used in the present study.

Interjudge reliability of the scoring procedure was evaluated by percentage of agreement between two judges: E, .86; I, .82; M, .84.

Fantasy Aggression. Fantasy aggression in the children was measured through an adaptation of the TAT procedure (Murray, 1943, pp. 3–5). A set of 10 pictures was designed. In each picture two boys are interacting. The pictures differed from one another in the degree to which the instigation to aggression was apparent.

To insure complete and accurate transcription of the stories, tape recordings were taken. An introductory period preceding the fantasy task served both to establish rapport between the child and the male examiner, and to familiarize the child with the recording device. Instructions were:

> I'm going to show you some pictures. These are pictures of two boys doing different things. What I'd like you to do is make up a story to each of these pictures. You can make up any story you wish; there are no right or wrong stories. Say what the boys are thinking and feeling and how the story will turn out.

The 10 pictures, in the order of presentation, were:

1. One boy is holding a basketball and the other boy is approaching him with arms outstretched.
2. One boy is stamping upon an ambiguous object and the other boy is reaching for the object.
3. One boy is sitting behind the other boy in a classroom and is leaning toward him.
4. One boy is walking down the street, and the other boy, with fists clenched, is glaring at him.
5. One boy, with fists clenched, is staring at the other boy who is sitting, head bowed, on a box.
6. One boy is sawing a piece of wood, and the other boy is leaning on a fence between them, talking to him.
7. The two boys, surrounded by a group of other boys, are approaching each other with arms upraised and fists clenched.
8. The two boys are making a fire. One boy is kneeling to arrange the wood and the other boy is approaching, ladened with wood for the fire.
9. One boy, who is looking back, is running down a street, and the other boy is running behind him.
10. Two boys are standing in a field. One boy, with his hand on the other boy's shoulder, is pointing off in the distance.

A fantasy aggression score was obtained for each *S* by counting the number of times the following acts appeared in his stories: fighting, injuring, killing, attacking, assaulting, torturing, bullying, getting angry, hating, breaking, smashing, burning, destroying, scorning, expressing contempt, expressing disdain, cursing, swearing, threatening, insulting, belittling, repudiating, ridiculing.

Fantasy aggression scores ranged from 1 to 15, with a mean of 5.3. The corrected matched-half reliability coefficient for the group of boys whose mothers encourage aggression (Low Anxiety group) was .88. For the group of boys whose mothers discourage aggression (High Anxiety group), the reliability coefficient was .85.

Interjudge scoring reliability coefficient was .92.

Overt Aggression. To measure overt aggression in the child, a modified sociometric device, the "Guess Who" technique, was adopted. The *Ss* were presented with a booklet containing a series of written descriptions of children, and asked to identify each of these descriptive characterizations by naming one or more classmates. Fifteen overt aggression items were used, such as "Here is someone who is always looking for a fight." A diversity of aggressive behaviors were included; items depicted verbal, unprovoked physical, provoked physical, outburst, and indirect forms of aggresive behavior.

An overt aggression score was obtained for each *S* by counting the number of times he was named by his classmates. There were substantial differences among the three classes in the distributions of the overt aggression scores; in order to combine into one distribution the scores of children in different classes, overt aggression raw scores were transformed into standard scores.

The biserial correlation coefficient between the overt aggression measure derived from the children and teacher entries for the same "Guess Who" aggression items was .76 ($p <$ .01).

Results

The intercorrelations among measures of extrapunitiveness, intropunitiveness, and impunitiveness (derived from the Picture-Frustration Study), fantasy aggression (derived from a modified TAT procedure), and overt aggression (measured by a near-sociometric device) are presented in Table 1. The intercorrelations are presented separately for the Low Anxiety group, the High Anxiety group and both groups combined.

All correlations are Pearson product-moment coefficients. Over-all correlations were obtained by pooling within group sums of cross products and sums of squares; this procedure is described by, for example, McNemar (1955). All probability values are based upon two-tailed tests of significance. Scatter plots revealed some skewness in a number of distributions; hence, all distributions were normalized for calculation of Pearson product-moment correlations.

Table 1

Intercorrelations Among Extrapunitiveness (E), Intrapunitiveness (I), Impunitiveness (M), Fantasy Aggression (FA), and Overt Aggression (OA) for Low Anxiety Group, High Anxiety Group, and Both Groups Combined

	E	I	M	FA	OA
E					
I	−.74** −.59** −.64**				
M	−.89** −.68**† −.78**	+.44* −.13† +.10			
FA	+.48* +.20 +.37**	−.38 +.11 −.14	−.44* −.39 −.40**		
OA	+.45* −.17† +.17	−.47* +.17† −.11	−.44* +.18† −.15	+.43* −.41‡ +.07	

Note.—In each set of three correlation coefficients, the first coefficient represents the correlation for the Low Anxiety group ($N = 23$), the second coefficient represents the correlation for the High Anxiety group ($N = 21$), and the third coefficient represents the combined correlation for both groups.

* $p < .05$ (two-tailed test).

** $p < .01$ (two-tailed test).

†, ‡ Low Anxiety group significantly different ($p < .05$, $< .01$, respectively) from High Anxiety group.

This study proposes that each measure of aggression employed constitutes a validating criterion for every other measure. For all seven correlations among the Picture-Frustration, TAT, and near-sociometric measures, the coefficients for the Low Anxiety group are greater (in a direction indicating greater validity of the measures) than for the High Anxiety group. In four of the seven comparisons, the difference between Low Anxiety and High Anxiety groups is significant beyond the .05 level of confidence.

The extent of the difference between Low Anxiety and High Anxiety groups is most apparent for the relationships of E, I, and M with overt aggression and for the relationship of fantasy aggression with overt aggression. The direction of difference in favor of the Low Anxiety group is maintained for the relationships of E, I, and M with fantasy aggression, but to a lesser degree. Thus, differences in construct validity which are dependent upon the population difference in degree of anxiety about aggression are clearly discernable for the relationships between nonovert (Picture-Frustration and TAT) and overt (near-sociometric) measures of aggression and are less discernable for relationships among the nonovert measures (Picture-Frustration and TAT) of aggression.

The intercorrelations among E, I, and M measures, which appear in Table 1, do not possess the same status as independent comparisons that the intercorrelations between either E, I, and M and the measures of overt and fantasy aggression possess. The closed nature of the scoring system for the Picture-Frustration responses automatically demands negative correlations among E, I, and M, since these three measures have only 2 *df*. That is, a high E score will automatically be accompanied by a low I score, a low M score, or both. Positive validity results require negative correlations between E and I and between E and M. Such negative correlations are demanded equally for both the Low Anxiety and High Anxiety groups. However, the negative correlations between E and I and between E and M in Table 1 are greater for the Low Anxiety group than for the High Anxiety group.

Discussion

When psychological studies produce discrepant findings, it has become routine to state that population differences may have produced the discrepancies. Differences in validity findings in studies employing identical measures are often vaguely explained in this manner. However, few attempts have been made to systematically define the psychological variables contributed by population differences and to empirically test the effects of these differences upon validity. In this study, population differences in degree of anxiety about aggression were shown to affect the construct validity results for a number of measures of aggression.

The general form of the conclusion drawn from this result is that there may be interactive effects between measures of personality variables and measures of psychological characteristics of the populations upon which the measures are obtained which influence other measures of the personality variable. Validity findings will be obscured by failure to consider such interactions.

The results of this study indicate that the validity of nonovert and overt measures of aggression can be more easily demonstrated for a group of boys who have low anxiety about aggression. The population difference in anxiety about aggression produces greater variability or inconsistency between nonovert and overt measures of aggression than among nonovert measures of aggression.

Certain limitations of the present study are apparent. The population difference in anxiety about aggression was measured indirectly by means of information concerning the inhibitory socialization attitudes and practices indicated by the mothers of the *S*s. Although there is substantial evidence in other researches that anxiety about aggression is produced by inhibitory socialization of aggression, measures of the *S*s' own responses which indicate anxiety about aggression would provide a more direct test of the hypothesis. In addition, the fact that the total group of boys employed in this study was significantly more extrapunitive and significantly less intropunitive and less impunitive than Rosenzweig's normative group makes the representativeness of the present sample suspect.

One other potential qualification was empirically assessed. An alternative explanation of the present results may be based upon differential attenuation in range of scores on the aggression measures for the Low Anxiety and High Anxiety groups. However, differences in standard deviations were not statistically significant for any aggression measure. It does not seem likely that the consistently lower correlations for the High Anxiety group were based upon greater attenuation in range of aggression scores for this group.

Summary

The results of this study suggest that it is inaccurate to describe measures of personality as valid or invalid in a general or overall sense. A single personality measure may possess different degrees of validity for different populations.

The hypothesis tested is that under conditions of high anxiety about aggression, the intercorrelations among various measures of aggression are significantly lower than under conditions of low anxiety about aggression.

Anxiety about aggression was measured by assessing the inhibitory socialization behavior of the mothers of 10- to 13-year-old boys.

Scores obtained from three techniques for measuring aggression in the boys (Rosenzweig's Picture-Frustration Study, a modified TAT, and a near-sociometric measure of overt aggression) were intercorrelated.

All intercorrelations among the measures of aggression were greater (in the direction indicating greater validity of the measures) for a group of boys with low anxiety about aggression than for a group of boys with high anxiety about aggression; for four of the seven intercorrelations, this difference was statistically significant. Also, correlations among the Picture-Frustration scores themselves indicated greater validity of these measures for the Low Anxiety group of boys than for the High Anxiety group.

The study demonstrates that, in validation research, more exact specification is required in the determination of conditions under which consistency among various aspects of an *S*'s responding may be expected and the conditions under which no consistency or inconsistency may be anticipated.

Received January 28, 1958.

REFERENCES

AMERICAN PSYCHOLOGICAL ASSOCIATION. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull.* (Suppl.), 1954, **51**, Part 2, 1–38.

CARLSON, V. R., & LAZARUS, R. S. A repetition of Meyer Williams' study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1953, **17**, 247–253.

DAVIDS, A. Some personality and intellectual correlates of intolerance of ambiguity. *J. abnorm. soc. Psychol.*, 1955, **51**, 415–420.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

DORKEN, H., & KRAL, V. A. The psychological differentiation of organic brain lesions and their localization by means of the Rorschach test. *Amer. J. Psychiat.*, 1952, **108**, 764–771.

FISHER, J., GONDA, T. A., & LITTLE, K. B. The Rorschach and central nervous system pathology: A cross-validation study. *Amer. J. Psychiat.*, 1955, **111**, 487–492.

FRENCH, ELIZABETH G. Interrelation among some measures of rigidity under stress and nonstress conditions. *J. abnorm. soc. Psychol.*, 1955, **51**, 114–118.

GRINKER, R. R., & SPIEGEL, J. P. *Men under stress.* New York: Blakiston, 1945.

HOLLENBERG, ELEANOR, & SPERRY, MARGARET. Some antecedents of aggression and effects of frustration in doll play. *Personality,* 1951, **1**, 32–43.

KUHLEN, R. G. *The psychology of adolescent development.* New York: Harper, 1952.

LESSER, G. S. Maternal attitudes and practices and the aggressive behavior of children. Unpublished doctoral dissertation, Yale Univer., 1952.

LURIA, A. R. *The nature of human conflicts.* New York: Liveright, 1932.

MAHL, G. The effect of chronic fear on the gastric secretion of HCl in dogs. *Psychosom. Med.*, 1949, **12**, 140–169.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1955.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus–response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

MURRAY, H. A. *Thematic Apperception Test Manual.* Cambridge: Harvard Univer. Press, 1943.

ROSENZWEIG, S., FLEMING, EDITH E., & ROSENZWEIG, LOUISE. The children's form of the Rosenzweig Picture-Frustration Study. *J. Psychol.*, 1948, **26**, 141–191.

WHITING, J. W. M., & CHILD, I. L. *Child training and personality.* New Haven: Yale Univer. Press, 1953.

WILLIAMS, M. An experimental study of intellectual control under stress and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21–29.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

EFFECTS OF VERBAL AND NONVERBAL REINFORCEMENT IN THE RORSCHACH

LEONARD R. GROSS[1]

Veterans Administration Hospital, Roanoke, Virginia

While there is some awareness of the dynamic interactions between examiner (E) and subject (S), the problem of just how and to what extent test results are influenced by these interactions have not been fully explored. It has been demonstrated that Rorschach responses can be influenced by orientational sets stemming from pretest practice (Leventhal, 1956), pretest suggestion (Abramson, 1951), and conscious instructions (Henry & Rotter, 1956), as well as by E differences (Sanders & Cleveland, 1953), but there are also reasons to believe that the E's actions throughout the testing situation may affect Rorschach performance (Wickes, 1956).

It would be desirable to show that different cues given by the E are responded to without conscious awareness by the S, resulting in a changed Rorschach protocol. This study attempted to reinforce general human content on the Rorschach, using the verbal reinforcer *good* and the nonverbal reinforcer *nodding*. The Ss were randomly selected psychiatric patients.

On the basis of previously mentioned studies it can be hypothesized that: (*a*) the verbal reinforcer *good* will increase the frequency of the reinforced responses over that of a control group; (*b*) the nonverbal reinforcer *nodding* will increase frequency of the reinforced responses over that of a control group; and (*c*) the verbal stimulus will be more effective than the nonverbal stimulus in increasing the reinforced responses.

Method

The Ss were selected from the psychiatric section of a university hospital on the following bases: (*a*) no history of organic brain damage, (*b*) a minimum of tenth grade education, and (*c*) no previous Rorschach experience. They were randomly selected from both the inpatient and outpatient services and placed in one of three groups in the following prescribed order: verbal reinforcement group (VR), nonverbal reinforcement group (NVR), or control group (C).

The Ss were excluded from the study if they did not meet both of the following criteria: (*a*) give at least one response involving general human content (humans, humanlike creatures, human anatomy) during the first two cards, and (*b*) give three responses per card for all ten cards.

Out of the 46 Ss tested, 6 did not meet the criterion of three responses per card, and 10 failed to produce one human response within the first two cards, leaving a total of 30 Ss, with 10 Ss in each group. The mean age of all the Ss was 34, with a range of 17 to 53. The sexes were evenly divided. The diagnoses were mixed and included neurotics, character disorders, and psychotics. The mean level of education for the 30 Ss was the twelfth grade, with 11 Ss having some college education. There were no significant differences in educational level among the three groups.

The Ss were presented with the complete set of Rorschach cards in the standard procedure for the free association with one variation in the Beck instructions, i.e., the inclusion of a sentence requesting three responses per card.

In the VR group, the E said "good" after each general human response. In the NVR

---

[1] The author wishes to acknowledge his indebtedness to W. J. Eichman and B. M. Smith of the Roanoke, Virginia, Veterans Administration Hospital for their guidance and assistance.

group, the *E* nodded his head once after each human response. In the C group, the cards were administered with the attempt not to offer any cues.

Posttest interviews revealed that none of the *S*s verbalized any awareness of the nature of the study.

Results

The mean number of general human responses were compared for the three groups. As a result of heterogeneity of variance, a square root transformation was applied to the raw scores of the individual *S*s. The means and variances of the transformed scores as well as the raw scores are found in Table 1. The analysis of variance of the transformed scores for the effect of reinforcement yielded an *F* significant at the .06 level. The groups were compared with each other by means of individual *t* tests. One-tailed *t* test was used as the direction was previously specified. The VR group gave more human responses than the C group at the .05 level of significance. The NVR group yielded more human responses than the C group at the .02 level of significance. The VR and NVR groups did not differ significantly from each other.

Table 1

Raw and Transformed Scores of the Three Groups

	Raw Scores		Square Root Transformation	
Group	Mean	Var.	Mean	Var.
VR	8.7	11.34	2.86	.32
NVR	10.3	27.12	3.09	.59
C	6.3	6.77	2.39	.33

Discussion

The results suggest that nodding or saying "good" will increase the frequency of preselected content responses in a Rorschach situation.

The first two hypotheses that the verbal reinforcer *good* and the nonverbal reinforcer *nodding* will increase the frequency of the reinforced responses appear to be substantiated. The third hypothesis that the verbal stimulus will be more effective than the nonverbal stimulus in increasing the reinforced response is rejected. The latter result is in conflict with previous studies using nonverbal cues (Taffel, 1955). A possible explanation is that the use of a flashing light as a nonverbal reinforcer in other studies was not perceived as part of the testing situation, while nodding was.

The most obvious implication of the results is that one cannot discount even minimal or unconscious cues of the *E* when analyzing a Rorschach protocol. It follows that interpretations of test responses and test behavior should not be considered separately but in light of the total situation. Both the *E*'s behavior and the *S*'s conception of the testing situation are of import.

While the results in this study are suggestive of some of the variables involved in the complicated interaction between tester and testee or interviewer and interviewee, it remains for future research to make experimentally clear other variables operating in such situations. It is probable that other aspects of Rorschach responses can be reinforced in a similar manner, but exactly which responses occur in sufficient number to be reinforced as well as how much of a variation in *E* behavior is necessary to produce a change in a response level has not been answered. Another question is just how important are the variables that can be affected. It seems that if classical scoring methods are used these variables can have considerable effect on the dynamic picture.

Summary

The main interest was interpersonal relations in a clinical situation. The study was designed to test the general hypothesis that *E-S* interaction is an important variable in test results. Thirty psychiatric patients were randomly selected and administered the free association of the Rorschach with the standard instructions modified so as to elicit three responses per card. The *S*s were then presented with either verbal reinforment *good,* nonverbal reinforcement *nodding,* or no reinforcement, whenever they gave a general human content response. It was found that both the VR and NVR groups gave significantly more of the reinforced responses than the C

Table 1

Simple Intercorrelation Among Scoring Categories
(From Wittenborn, 1950a)

No.	Symbol	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	*W*																		
2	*D*	−194																	
3	*d*	−228	669																
4	*Dd*	−138	608	749															
5	*S*	−143	464	512	660														
6	*M*	−041	498	560	665	616													
7	*FM*	−058	625	429	381	288	359												
8	*m*	115	403	351	297	287	379	375											
9	*K*	255	330	246	259	155	269	281	425										
10	*FK*	203	275	238	221	232	154	201	240	411									
11	*F*	−052	693	705	762	635	438	285	202	147	198								
12	*Fc*	100	587	583	605	519	579	326	450	349	153	460							
13	*c*	144	138	116	108	097	113	136	182	310	193	108	191						
14	*C′*	016	466	510	503	508	532	345	468	310	114	377	546	076					
15	*FC*	023	541	429	457	374	342	361	563	181	028	305	571	−084	464				
16	*CF*	410	241	090	114	−008	−108	−025	113	353	285	179	117	251	089	133			
17	*C*	170	178	072	137	066	−020	075	098	259	270	039	098	261	109	177	235		
18	*P*	−095	208	266	163	034	177	221	363	198	−005	025	357	187	206	356	−033	216	
19	*O*	−011	436	570	579	565	624	303	456	305	135	415	646	181	379	550	046	172	659

Table 2

Partial Intercorrelations of Rorschach Scoring Categories
(*R* Held Constant)

No.	Symbol	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	*W*																		
2	*D*	−582																	
3	*d*	−486	−064																
4	*Dd*	−370	−383	309															
5	*S*	−250	−163	088	380														
6	*M*	−128	−150	139	341	386													
7	*FM*	−138	295	−051	−192	−090	−009												
8	*m*	081	−041	−042	−180	001	114	135											
9	*K*	239	−109	−160	−171	−141	−001	044	277										
10	*FK*	183	−132	−103	−162	007	−117	−025	073	297									
11	*F*	−210	−028	215	324	326	−134	−362	−337	−358	−186								
12	*Fc*	049	−110	062	068	167	241	−152	179	069	−178	−249							
13	*c*	129	−071	−061	−089	−025	−013	029	103	255	131	−079	077						
14	*C'*	−042	−045	152	113	268	285	037	280	098	−123	−123	264	−038					
15	*FC*	−038	072	−055	−038	037	−040	033	398	−095	−257	−327	273	−248	210				
16	*CF*	404	−080	−266	−257	−258	−415	−269	−046	254	189	−126	−159	204	−107	−066			
17	*C*	154	−041	−170	−079	−087	−212	−069	−011	185	205	−238	−089	227	−017	058	177		
18	*P*	−115	092	195	017	−099	077	140	316	132	−084	−211	324	156	125	309	−099	182	
19	*O*	−075	−121	268	270	353	427	−025	263	082	−096	−048	434	089	108	337	−162	060	678

Table 3

Simple Intercorrelations Among Rorschach Scoring Categories
(From Wittenborn, 1950b)

No.	Symbol	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	*W*																			
2	*D*	342																		
3	*d*	191	638																	
4	*Dd*	084	410	655																
5	*S*	261	339	305	344															
6	*M*	281	437	270	281	179														
7	*FM*	282	513	270	318	060	480													
8	*m*	296	306	395	512	245	135	294												
9	*k*	140	222	187	105	347	011	−098	131											
10	*K*	470	507	519	413	234	245	299	267	073										
11	*FK*	239	313	239	357	319	378	196	181	123	275									
12	*F*	349	727	741	693	483	129	248	421	222	505	230								
13	*Fc*	076	249	279	302	196	146	239	293	013	129	199	179							
14	*c*	437	424	545	403	286	125	102	363	160	458	187	475	100						
15	*C′*	337	481	508	323	179	277	241	219	059	380	260	354	165	453					
16	*FC*	335	588	378	313	292	342	436	309	176	224	278	356	289	389	430				
17	*CF*	513	509	592	345	250	225	147	153	283	507	203	464	105	549	581	401			
18	*C*	419	249	304	189	104	058	019	331	047	235	014	319	034	445	149	044	261		
19	*P*	063	096	000	−125	075	094	029	−093	−004	−039	109	003	−031	006	008	031	−106	009	
20	*O*	210	211	135	119	484	085	123	275	046	155	113	300	120	200	135	349	018	−006	235

Table 4

Partial Intercorrelations Among Rorschach Scoring Categories
(*R* Held Constant)

No.	Symbol	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	*W*																			
2	*D*	−313																		
3	*d*	−394	−111																	
4	*Dd*	−400	−435	329																
5	*S*	−001	−227	−126	041															
6	*M*	069	138	−100	006	−041														
7	*FM*	026	189	−191	−002	−242	342													
8	*m*	064	−248	068	312	021	−081	083												
9	*k*	018	032	011	−063	273	−103	−253	024											
10	*K*	218	−071	107	035	−094	−021	−001	−023	−091										
11	*FK*	068	018	−043	185	181	270	030	023	045	086									
12	*F*	−270	−162	241	349	128	−546	−421	037	037	−056	−155								
13	*Fc*	−090	007	105	165	069	030	121	192	−057	−054	113	−134							
14	*c*	189	−207	199	052	006	−162	−254	132	031	168	−016	−056	−077						
15	*C'*	070	010	173	−042	−120	058	037	−046	−083	076	091	−281	016	205					
16	*FC*	022	153	−159	−123	−006	117	201	040	045	−222	092	−423	156	065	155				
17	*CF*	265	−151	206	−116	−097	−072	−255	−215	177	193	−026	−245	−100	285	363	024			
18	*C*	289	−151	050	−060	−090	−114	−195	199	−041	023	−127	015	−075	310	−059	−231	043		
19	*P*	061	164	−025	−182	075	094	022	−116	−009	−065	109	−030	−039	−007	−004	023	−155	001	
20	*O*	067	−104	−144	−096	408	−048	−026	163	−025	−030	012	095	041	041	−030	227	−232	−124	240

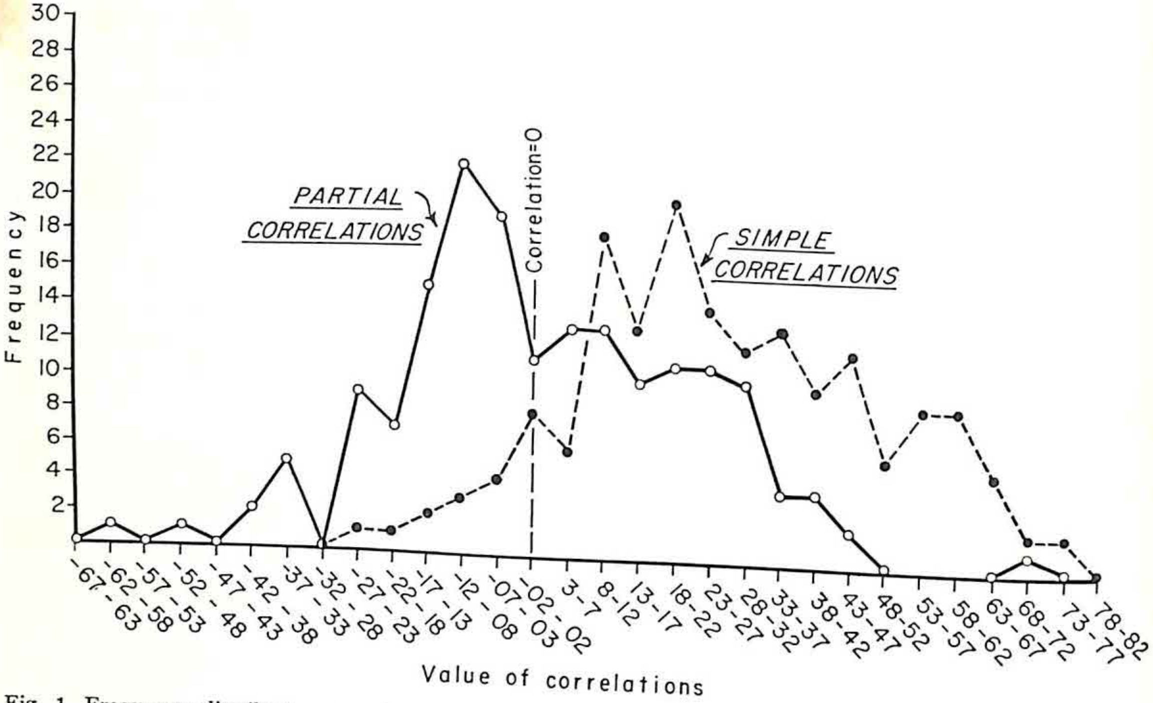


Fig. 1. Frequency distribution correlations before and after partialling out number of responses. Data from Wittenborn (1950a).

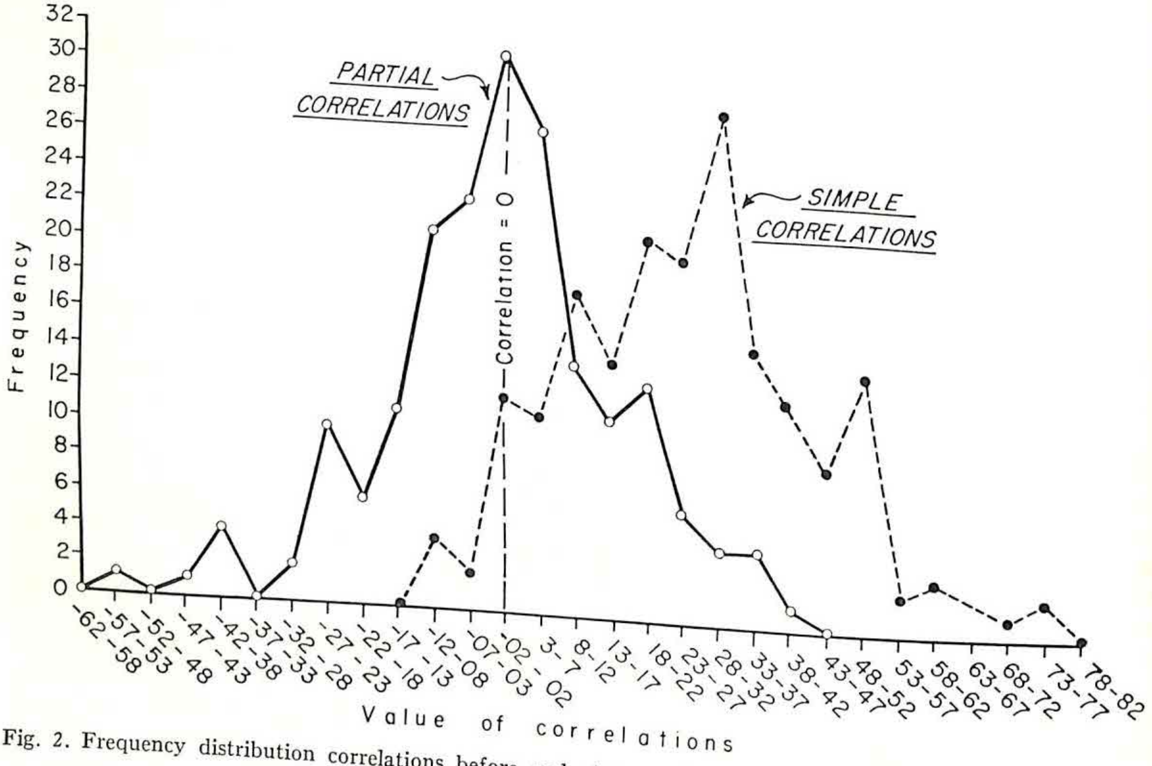


Fig. 2. Frequency distribution correlations before and after partialling out number of responses. Data from Wittenborn (1950b).

− 1.00. This effect becomes less important the greater the number of mutually exclusive categories intercorrelated, which can be seen in the present matrices of partial intercorrelation (Tables 2 and 4). The pooled partial correlations between the five location categories in the two matrices have a median value of − .195, with 15 of these 20 values negative. A sign test shows this negative tendency to be significant at beyond the .025 level. The Rorschach determinant scores, in which the categories are mutually exclusive but more numerous, have a median intercorrelation of − .001, and these partials are not significantly negative. This problem of deflation of correlations by mutually exclusive categories is, of course, not resolved by the procedure of partial correlation.

These considerations severely limit any confidence in the resultant factor structure. The general inflation of the entire matrix by a variable R would be reflected in spuriously high communalities. The mutually exclusive categories would artificially decrease the correlations among some variables and not others.

Received February 10, 1958.

REFERENCES

Simon, H. A. Spurious correlation: A causal interpretation. *J. Amer. statis. Ass.*, 1954, **49**, 467–479.

Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950a, **14**, 261–267.

Wittenborn, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.*, 1950b, **14**, 469–472.

Yule, G. U., & Kendall, M. G. *An introduction to the theory of statistics.* (11th ed.) London: Charles Griffin, 1937.

Zeisel, H. *Say it with figures.* New York: Harper, 1947.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

SOME COMMENTS ON CONFOUNDED CORRELATIONS AMONG RORSCHACH SCORES

J. R. WITTENBORN
Rutgers University

Glickstein has indicated that he believes that R, the total productivity score on the Rorschach, influences the correlation between other Rorschach scores in the manner of a common third variable (Glickstein, 1959). He then proceeds to examine the way in which this putative common third variable affects intercorrelations among Rorschach scoring categories; he does this by holding the variance of R constant through the device of partial correlation. He finds that a substantial portion of the resulting partial correlations carries a negative sign. He then ascribes the preponderance of negative partial correlations to the fact that Rorschach scoring procedures involve the use of a "small" number of mutually exclusive categories. After assuming that R inflates all the intercorrelations in the manner of a common third variable and after further assuming that the intercorrelations among the scoring categories are simultaneously suppressed by the mutually exclusive qualities inherent in the Rorschach scoring procedure, he concludes that confidence in the result of factor analytic studies of Rorschach scoring categories must be severely restricted.

The writer's views concerning the nature and significance of the R variable are different from Glickstein's. The number of responses in the various scoring categories determines R, and R does not delimit the number of responses in any scoring category; accordingly, the writer does not consider it useful to regard *R* as a common third variable. *R* is generated by the number of responses in the various scoring categories, and the writer shares the interpretation of Coan (1956) and of Stotsky (1957) who see the relationship between any response category score and the total productivity score, *R*, as a part-whole relationship. With this interpretation, partialling *R* from the correlations between Rorschach scoring categories would be something like partialling the total score on a test from the correlations between the subtest scores or like partialling the full scale score on the WAIS from the intercorrelations among the various subtests. The writer does not have any questions about the Rorschach or any other test which require this kind of statistical exercise, but he suspects that such a partial correlation procedure would yield many negative partial intercorrelations among the WAIS subtests, just as it was found to yield negative partial intercorrelations among Rorschach categories.

The effect of the mutually exclusive feature of the Rorschach scoring procedure need not be as great as Glickstein imagines. If the correlations were based on a sample of persons who gave but one response each, there would have to be an inverse relationship between any two scoring categories for that response, and as Glickstein notes, if only two scoring categories were available (instead of the numerous determinant and location categories that are available), there would also have to be a negative relationship between the two scoring categories. Because of the diversity of available determinant and location categories and because most subjects give numerous responses, it is not necessary for the intercorrelations between the categories to be small or negative, even in samples where *R* is held constant, e.g., in hypothetical samples comprising persons giving the same number of responses. Since the variance of *R* is generated by the variance of the scoring categories, it is not surprising that the correlations between the scoring categories are greatly reduced when the variance of *R* is eliminated from the correlations by the partial correlation procedure. In view of this, it would be difficult to ascribe any general significance to the fact that many of Glickstein's partial correlations are small and negative.

The writer's interest in the Rorschach method was generated by clinical use, and during the 1940's several studies were undertaken for the purpose of exploring some of the assumptions which were explicit in the method or were inferable from standard procedures (Wittenborn & Sarason, 1949; Wittenborn: 1949a, 1949b, 1949c, 1950a, 1950b, 1950c; Wittenborn & Holzberg, 1951). At that time it was recognized that the nature and the distribution of some of the variables do not always meet the exact requirements of some of the statistical procedures employed, e.g., "Since the present investigation is concerned with the conceptual implications of the Rorschach *scoring procedures*, it was decided to study the scores directly with whatever imperfections they may possess and not attempt statistically to refine them or their intercorrelations" (Wittenborn, 1950a, pp. 263–264).

When the study from which Glickstein draws his data was planned, it was hypothesized that the scores would be intercorrelated in such a way as to generate a general factor of productivity. Total number of responses, *R*, was left in the matrix in order to identify a factor of productivity and indirectly to facilitate the identification of other factors. The correlations did not sustain the hypothesis of such a general factor. Nevertheless, a group factor of productivity was found, and a similar factor has been described by other investigators using other samples (Coan, 1956; Williams & Lawrence: 1953, 1954). The fact that *R*, the total productivity score, bears a part-whole relationship with various Rorschach scores has been recognized by most investigators, and some of them have eliminated this distorting effect of *R* on factor patterns by eliminating *R* from the total matrix of correlations (Coan, 1956; Stotsky, 1957). Both Coan and Stotsky explicitly recognize that the nature of the Rorschach procedure per se generates varieties of confounding among the Rorschach variables, and they have been resourceful in attempting to control these effects in their analyses.

To describe some of the intercorrelations between Rorschach variables as spurious is correct from the standpoint of the statistical denotation of the word spurious, but so naming this effect does not reduce it or make the various sources of confounding any less real or make R a less useful part of the Rorschach method of personality analysis. At present, one has the unhappy choice of studying the Rorschach "as it is," of studying it "as it isn't," or of ignoring it altogether. Perhaps someone will soon devise a new procedure which will incorporate the merits of the Rorschach and eliminate some of the intricacies and confounding which make its proper statistical study so forbidding.

Received October 20, 1958.

REFERENCES

Coan, R. A factor analysis of Rorschach determinants. *J. proj. Tech.*, 1956, **20**, 280–287.

Consalvi, C., & Canter, A. Rorschach scores as a function of four factors. *J. consult. Psychol.*, 1957, **21**, 47–51.

Glickstein, M. A note on Wittenborn's factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1959, **23**, 69–77.

Stotsky, B. A. Factor analysis of Rorschach scores of schizophrenics. *J. clin. Psychol.*, 1957, **13**, 275–278.

Williams, H. L., & Lawrence, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.*, 1953, **17**, 261–264.

Williams, H. L., & Lawrence, J. F. Comparison of the Rorschach and MMPI by means of factor analysis. *J. consult. Psychol.*, 1954, **18**, 193–197.

Wittenborn, J. R. Certain Rorschach response categories and mental abilities. *J. appl. Psychol.*, 1949a, **33**, 330–338.

Wittenborn, J. R. A factor analysis of discrete responses to the Rorschach ink blots. *J. consult. Psychol.*, 1949b, **13**, 335–340.

Wittenborn, J. R. Statistical tests of certain Rorschach assumptions: Analyses of discrete responses. *J. consult. Psychol.*, 1949c, **13**, 257–267.

Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950a, **14**, 261–267.

Wittenborn, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.*, 1950b, **14**, 469–472.

Wittenborn, J. R. Statistical tests of certain Rorschach assumptions: The internal consistency of scoring categories. *J. consult. Psychol.*, 1950c, **14**, 1–19.

Wittenborn, J. R., & Holzberg, J. D. The Rorschach and descriptive diagnosis. *J. consult. Psychol.*, 1951, **15**, 460–463.

Wittenborn, J. R., & Sarason, S. B. Exceptions to certain Rorschach criteria of pathology. *J. consult. Psychol.*, 1949, **13**, 21–26.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

THE SELECTION OF THE MENTALLY DEFICIENT FOR VOCATIONAL TRAINING AND THE EFFECT OF THIS TRAINING ON VOCATIONAL SUCCESS[1]

LAWRENCE COWAN AND MORTON GOLDMAN
University of Kansas City

The problems of the mentally deficient have recently undergone a period of extensive examination. One of these problems that is receiving the most stress at this time is in the area of vocational adjustment (Engel, 1950). The present study examines the relationship of training to this vocational adjustment and explores several criteria that could possibly be used to predict the success or failure of the mentally deficient in achieving this adjustment.

Some authors (Engel, 1950; Keys & Nathan, 1932) have expressed the opinion that training is of questionable value for the mentally deficient unless it is specially focused toward a definite job. However, the sheltered workshops available in some communities, giving a more global type training with awareness of factors dealing with personality adjustment, have done a great service for the mentally deficient in affording them an opportunity to receive training and helping them to achieve a moderate degree of vocational success (Gellman, Gendal, Glaser, Friedman, & Neff, 1956; Weinberg, 1955). As a source of training, these agencies are limited by the community's effort to expand them. At the present time they are quite inadequate to handle the number of mentally retarded and deficient who are in need of vocational training (Gellman et al., 1956).

By utilizing training cases from other than workshop sources, it will be the intention of this study to examine whether these other places of training can adequately prepare the mentally deficient for eventual job placement. If this is so, then the community will have other resources to expand in helping the mentally retarded and deficient to receive vocational training.

The first hypothesis (H_1) this study examines is: The mentally deficient, as a result of vocational training, will be able to get and keep a job more successfully than a matched nontrained group. This hypothesis may seem somewhat naive, since one is inclined to think that any trained person would be more successful than one not trained in getting and keeping a job. But this is not necessarily so, for at one time it was thought that the jobs a mentally deficient person could do were so simple as to warrant no training, and, if training were given, it would not appreciably benefit that individual (Engel, 1950).

The second hypothesis (H_2) that will be examined is: The ability to get and keep a job for the mentally deficient (IQ level 40–79) varies directly as the IQ level.

The concepts of intelligence and IQ have of necessity undergone many changes in emphasis during the past years, swinging from no credence placed in the terms to attempting to adjust an individual's educational and vocational development by using the "intelligence quotient." Fortunately for many, the use of the IQ score has shifted back toward the middle of these extremes. There is one area, however, that has been noticeably slow in using a more operational concept of intelligence—this has been in the vocational counseling of the mentally retarded and deficient.

---

[1] This paper is a condensation of a dissertation (Cowan, 1957) submitted to the Graduate School of the University of Kansas City by the first author. The second author served as advisor.

It is here, perhaps more than anywhere else, that the IQ score has been "glorified." It should be pointed out that most professional books and articles in the field of vocational counseling of the mentally retarded warn against adopting the IQ score as categorically placing the individual into a class of potential achievement (Mausner, 1953). This warning is continually being voiced due to the large incidence of workers doing just that. It is common to find vocational counselors with the opinion that given two mentally deficient individuals, one with a tested IQ of 55 and the other with an IQ of 75, the one with the higher IQ will make a better vocational adjustment due to the difference in IQ points!

The third hypothesis (H_3) is concerned with the examination of several other factors which are purported to play a direct role in predicting vocational success. It states: Vocational success can be predicted from effort made in attempting to locate a job, formal education, and past work experience. A vocational rehabilitation counselor must make a judgment as to the potential of the mentally deficient client prior to accepting him for services. In doing this he always uses the information available. It seems reasonable that a counselor might consider a person who has made little or no attempt to find a job as not having the vocational potential that one who has exerted more effort will have. Further, the counselor might think that the more education achieved, the better the chance for success. Also, he might think that if the individual has some work experience, he will make a better adjustment to training on a new job. The work history is probably the one factor that is most important to the vocational counselor, for it is here that the counselor feels he can best discover the adaptability of his client. All these factors taken as a whole give the counselor an opinion as to the potential of the individual for vocational success. The IQ score and the case history, which give him information of the nature stated in the third hypothesis, are used to arrive at a decision as to whether there is a "reasonable expectation" that the individual can be vocationally rehabilitated. It is true that these factors of IQ, work history, etc. are never taken separately but are examined in relationship to the total picture of the individual. It is the authors' view that, regardless of the attempt toward a total picture, this picture will not yield reliable information because the criteria used may not be valid. If the value of training of the mentally deficient is to be judged by vocational counselors, it might be necessary for them to re-evaluate their criteria. It may be that techniques of making decisions which relate to the advisability of training the mentally deficient will need to be adjusted drastically as more up-to-date information becomes available.

Method

As a first step, the files of a district office of vocational rehabilitation[2] were sorted for all the mentally deficient clients given training since the agency opened its office in 1952. These individuals constituted the experimental group. Those persons who were referred with mental deficiency plus a physical handicap were ruled out since the physical aspects of their disability would contaminate the results of the study. When this sorting was complete, there were 22 individuals who met the requirements of the experimental group, that is, mentally deficient (congenital defect) with no accompanying physical disability, who received vocational training. Twenty were available for the complete work-up; the other two had moved from the city after they had received their training. The borderline upper limit of retardation or deficiency in IQ as stated by Wechsler (1944) was used as an upper cutoff score, that is, an IQ of 79. Since no individual with an IQ of less than 40 was accepted for training, this constituted the lower limit.

A questionnaire was developed to use in the follow-up interview of this study. It was designed to yield the following information: name, age, sex, race, education, referral source, IQ, date of IQ testing, date of contact, type of training, where training was given, and for how long. If employed since the time of training, information was obtained as to what type of employment, for how long the job was held, and how the job was obtained.

---

[2] The Office of Vocational Rehabilitation, Kansas City, Missouri.

When this information was gathered on all 20 cases, a comparable group of mentally deficient individuals was taken from the files of a private rehabilitation agency to correspond to the experimental group. These people were selected on the same criteria as the first group from a total of 53 cases. Twenty people were selected who matched the experimental group as closely as possible on sex, age, education, race, IQ, and past employment. The major difference between these groups was that the second group (the control group) had received no vocational training, whereas the first group (the experimental group) had received vocational training. These individuals in the control group were referred to the private rehabilitation center for IQ tests.[3] These same people had never heard of the Office of Vocational Rehabilitation, and they were asked if they would have gone to Vocational Rehabilitation had they known about the services available there. Their answers were in the affirmative. It is thus assumed that the difference in referral source between the control group and the experimental group is one of lack of knowledge concerning the services available, and that the placement in the experimental or control group was a result of chance factors operating in the environment.

Following the selection of each group, all the members of the groups were interviewed personally, and the questionnaire was administered. Also, the mothers of all the subjects (*S*s) were interviewed in order to avoid the possibility that the individuals with higher IQs could give more complete information.

The results obtained for the experimental and control groups on the matching variables were as follows: There were 15 males and 4 females in the experimental group and 14 males and 4 females in the control group. All the *S*s in the study were white. The mean age, IQ, and grade for the experimental group were 20.8, 64.7, and 4.15; the corresponding scores, in the same order, for the control group were 22.9, 61.2, and 5.15. The differences between the standard deviations of these corresponding measures were small. The mean time that elapsed between time of testing and contact for the experimental group was 3.1 years and for the control group was 2.7. And finally, the number of cases with no past work history was 12 for the experimental group and the same number for the control group.

The 40 cases (20 in control and 20 in experimental) were interviewed to discover what had transpired since the time of training for the experimental group and since the time of testing for the control group. Information was obtained as to which individuals had secured employment and how long they had remained employed, whether those employed of the trained group had found jobs in the area of their training, and whether they had been successful in making a vocational adjustment. An adequate vocational adjustment was said to exist if the individual had held a paying position for at least 12 months. For the purposes of this study, a paying position consisted of employment where the individual was receiving at least the minimum wage set by the laws of the state[4] and was working enough hours during the week to result in a salary that was satisfactory to him. It should have afforded him a sufficient income to care adequately for his needs.

Results

After all questionnaires were completed, the data was combined and yielded the following over-all information: Of the 20 experimental cases (those receiving vocational training), 12 were employed at the time of contact and had been in that status for at least one year. They were receiving an adequate income as defined in the preceding paragraph. All 12 of these cases were placed in their jobs by the vocational counselor.[5] The successful group consisted of 9 males and 3 females which corresponds identically with the three to one ratio of males to females in the total group (15 males, 5 females). The mean age of the successful group was 21.75 years which does not

---

[3] The question can be raised whether all of the control group would have been acceptable to the Office of Vocational Rehabilitation for vocational training. When the records of the office used in this study were examined, they showed that the total number of mentally retarded and deficient referrals was 28. Since 22 of them were accepted, it is assumed that the majority of the control group would also have been accepted.

[4] In this case, $1.00 per hour.

[5] The implications of this will be discussed below.

Table 1

Effect of Training on Vocational Success

Status	Trained	Untrained	Total
Unemployed	8	16	24
Employed	12	4	16
Total	20	20	40

Note.—$\chi^2 = 8.437$ (corrected for continuity); $p < .001$.

vary significantly from the total mean age of the group (20.8 years).

The same criteria for assessing vocational success were applied to the control group (those receiving no training). Of these 20 cases, 4 had achieved a successful vocational adjustment. To test the significance of the relationship of training and employment, the chi-square test (corrected for continuity) was used.[6] The results of this test are shown in Table 1.

It can be seen from Table 1 that the probability of these results occurring by chance alone is less than one out of a thousand. These results were considered to be highly significant, and H_1 is supported.

To test H_2, the mean IQ of those individuals who achieved a successful vocational adjustment in both groups was compared with the mean IQ of those in their respective groups who were unsuccessful. Table 2 shows the mean and standard deviation of the IQs of the successful and unsuccessful *S*s in each group.

When the means were submitted to the *t* test, it was discovered that the difference between the mean IQs of the successfully trained and the unsuccessfully trained did not vary sufficiently to warrant the statement that they were from different populations. Likewise, the difference between the two means of the untrained was not sufficient to state they were from different populations. In other words, the differences of IQ between the successful and unsuccessful in either the trained or untrained group were not statistically significant. Further, the differences between the successful in both groups as compared to the unsuccessful in both groups were not significant. These findings then do not support H_2.

Table 2

IQ of Successful and Unsuccessful Subjects

Subjects	$\bar{X}$	*n*	*s*
IQ of successful, trained	63.9	12	9.2
IQ of unsuccessful, trained	66.0	8	9.0
IQ of successful, untrained	61.0	4	7.7
IQ of unsuccessful, untrained	61.2	16	8.4

In the statistical examination of the third hypothesis, it was necessary to test each of the three factors separately. The first factor examined was formal education. Table 3 shows the mean level of education in terms of grades and the standard deviation for the groups.

The means were submitted to the *t* test and the findings were: The difference between the means of educational level of the successfully and the unsuccessfully trained was not significant at the .05 level. The same results were found for the difference between the mean level of education of the untrained group. The difference between the means of the successful and unsuccessful *S*s, irrespective of training, was submitted to the *t* test, and the difference was also not significant at the .05 level.

The factor effort exerted to find employment was examined next. One point was given to an *S* for each attempt made to find employment. Comparing the median effort exerted by each group showed that the trained group had a median effort of 1.09, and the untrained group had a median effort of .50. The median test was used to test for significance. It showed that the difference in effort

Table 3

Level of Education for the Successful and Unsuccessful

Subjects	$\bar{X}$	*n*	*s*
Level of education of successful, trained	4.92	12	4.1
Level of education of unsuccessful, trained	3.00	8	3.4
Level of education of successful, untrained	6.75	4	3.9
Level of education of unsuccessful, untrained	4.75	16	4.7

---

[6] For all statistical methods used in this study, the .05 level of confidence will constitute significance.

between the trained group and the untrained was highly significant ($p < .001$). However, it must be pointed out that to receive training it was necessary for the trained group to go to the Office of Vocational Rehabilitation. This automatically gave them one point for effort. Since the combined median of both groups was .93, all the *Ss* in Group I were over the median. This accounted for the high significance in the trained group. This statistical test thus supports the aspect of H_3 that states effort is indicative of vocational success.

The third factor of H_3 examined was the relationship of work history to vocational success. Since the two groups were matched for the number having previous work experience, the support of H_1 established the fact that work history was not pertinent. Each group had 12 *Ss* who had past work experience. What remained to be explored was whether those individuals in each group who achieved vocational success consisted of a statistically significantly higher number who had had past experience. In each group, the number of persons with past experience was identical to the number without. For the 12 successfully trained, 6 had had past experience and 6 had not. These results indicated that there was no difference between having had or not having had past work experience and achieving vocational success. This finding, then, did not support the third factor of H_3.

The final result of the study examined was the length of time between the time of IQ testing and contact. It already has been stated that this time for each group was averaged, yielding a mean elapsed time for the training group of 3 years, 1 month. The elapsed time for the control group was 2 years, 7 months. If the average time spent in training for Group I (4.8 months) is subtracted from the average time elapsed for this group, a corrected figure of 2 years, 9.2 months is reached. This constitutes the time available for Group I to find employment. When these figures were compared to the mean time elapsed for Group II to find employment (2 years, 7 months), it was found that both groups had approximately the same length of time available to achieve vocational success as defined here.

Discussion

The primary concern of this study is the relationship of training to the vocational success of the mentally deficient. The results of the statistical analyses in the preceding section indicates that those receiving training made a better vocational adjustment than those who did not. When the variations of IQ, age, education, etc. were controlled, the results were still the same. Training made the difference. But training to the mentally deficient means more to them than just an opportunity to learn new skills. It means that someone is interested in them, someone is there to encourage their efforts and to help them handle the disappointments and frustrations that arise. They may not have received this support at home. Under a healthier atmosphere the mentally deficient, as well as anyone, can and will learn more and produce more. This brings us to the subject of the home life of the mentally deficient and how it relates to vocational adjustment. Talking with the mothers of these 40 cases brought to light some interesting differences between the two groups. These differences were not objectively measured in this study, but they manifested themselves too frequently to be ignored. In talking to the mothers of the cases in the control group (no training), it was commonplace to hear them say, "John can't work, he's not bright enough. Everyone would take advantage of him. He couldn't get along by himself—he's much better off here at home where I can watch him. Oh, yes! He does a lot of work around the house." This quotation is a composite of remarks from several mothers, but it represents the general feelings held by the mothers of *Ss* in this group. This attitude on the part of the mother may have come after John showed no inclination for outside work and was used defensively by the mother. It may have come before John was old enough to work and represented the protective shield that had surrounded him all his life. Either way, however, he is receiving no encouragement to accomplish on his own or to prepare himself for the time when he may be on his own. It was striking to compare these attitudes with those of the mothers of the group that received training. These mothers, although they had the same fears as the other group, were working as much as possible for the rehabilitation of their sons and daughters. This attitude might have become manifest due to the abilities shown by the mentally deficient in their training, but it may be more indicative of the nature of the environment in which they were raised. Perhaps the attitude of helpfulness and encouragement results from parental ability to accept the mentally deficient as an individual and implies no attempt to push him beyond his capacities nor to shelter him to such an extent that his capabilities are never discovered. Much has been said in the literature about the need to accept the physically and mentally disabled, but how does one accept a son or daughter who is unconsciously being rejected? The answer of course is that the individual is not accepted. There is more than a razor's edge between intellectual understanding and emotional acceptance.

The foregoing constitutes an attempt to understand some of the problems of the mentally deficient. Application of these thoughts to the experiment reported on here, however, can lead to some practical observations. The trained group did achieve a better vocational adjustment than the nontrained. Why this is so is still not too clear. Are those selected for training in a better position to profit from this emotionally, or does the training itself serve as a vehicle for the expression of undeveloped abilities? This question leads to need for more experimental evidence to establish the relationship of emotional development to vocational and social success. From the results of this study, it appears that many mentally deficient individuals would profit from vocational training. This training should of course be adjusted to fit the abilities and requirements of the individual. The individuals who had a healthy psychological home life could benefit from the combined effects of the encouraging atmosphere plus the new skills acquired. The mentally deficient are intellectually immature and, as is often the case, emotionally immature as well. They are more dependent upon their parents, and if a training program is to be effective, the agency responsible for the training may also have to work with the parents.

For along time now the relationship of tested IQ to job success has been known to be uncorrelated (Brainerd, 1954). The IQ test is an important tool in psychology, as well as in the field of education and counseling, with the mentally deficient. However, using IQ scores as a predictive device for vocational success is not supported by the evidence in the present study.

The formal education of the mentally deficient should also be given some attention. Like the IQ score, its use as a predictive device for vocational success is not supported by the evidence in this study.

Measuring the effort exerted by the *Ss* to find employment produced some problems. It was intended in using this measure to tap all methods of securing employment. One of the problems, thus, was to decide whether *Ss* in Group I (the experimental group), who were all trained and placed through the Office of Vocational Rehabilitation, should receive credit. Strictly speaking, making contact at the Office of Vocational Rehabilitation was an attempt to secure employment, and therefore the *Ss* in Group I all were allotted at least one point. However, the question can be raised as to whether this method of obtaining employment was comparable to the other methods of finding a job used by the control group, Group II. In one respect the individual assumed much more responsibility by cooperating with the plans formulated with the vocational counselor, but he also does not have the problem of finding a job opening on his own. This was done for him by the vocational counselor or the placement service. Because of these difficulties in measuring the effort exerted in finding a job, it cannot be said that a relationship has been found in this study between the effort exerted in locating a job and ultimate vocational success. It might even be possible to assume that it was the job placement that accounted for the success of the trained group rather than training.[7]

In the experimental group, the group that received training, we note that 8 of the 20 still remained unemployed. Our data shows that these "failure" cases cannot be distinguished from the "successful" cases by IQ, age, previous education, sex, or previous work history. How then can we account for these "failure" cases? The answer to this question will, of course, demand much additional research. The study reported on here had too few cases and was not designed to explore this area. However, certain suggestions can be made. One area that might be examined is the personality of the trainee in combination with the personality of the instructor, and the relationship established between the instructor and trainee. Another factor might be the type of training given and the demand that exists for people with this type of training. And finally, of high importance would be the effort and ability of the placement counselor who works with the trainee trying to locate a position for him.

---

[7] A study is presently under way which will compare two groups of *Ss*—one group given placement aid and training, the other group given only placement aid.

Summary

This study dealt with the relationship of training to the vocational success of the mentally deficient. Twenty cases of mental deficiency with no accompanying physical disability who received vocational training from the State Department of Vocational Rehabilitation were matched with 20 nontrained mentally deficient individuals for age, sex, race, IQ, education, past work experience, and elapsed time since the time of their IQ testing.

Administering a questionnaire, it was discovered that the trained group had a significantly larger number of vocationally successful individuals than the nontrained group. Further, this success was unrelated to their IQ level, formal education, or past work experience. Since the trained group was placed in their jobs by a vocational counselor, the results may be influenced by the placement itself, which is an integral part of rehabilitation. The measurement of effort scale showed several weaknesses in operation, and no assumptions concerning the relationship between effort and job success seemed feasible.

Received February 10, 1958.

REFERENCES

Brainerd, B. R. Increasing job potential for the mentally retarded. *J. Rehab.*, 1954, **20**, 4–6.

Cowan, L. The relationship of training to the vocational adjustment of the mentally deficient. Unpublished master's dissertation, Univer. of Kansas City, 1957.

Engel, Anna M. Employment of the mentally retarded. In S. G. DiMichael (Ed.), *Vocational rehabilitation of the mentally retarded*. Washington, D. C.: U. S. Government Printing Office, 1950. Pp. 80–107.

Gellman, W., Gendel, H., Glaser, N., Friedman, S., & Neff, W. Adjusting people to work. Unpublished project of Jewish Vocational Service and Employment Center, Chicago, Ill., 1956.

Keys, N., & Nathan, Jeannette M. Occupations for the mentally handicapped. *J. appl. Psychol.*, 1932, **16**, 497–511.

Mausner, H. The qualitative use of aptitude testing with the handicapped. Paper read at the Jewish Vocational Service Midwest Regional Conference Proceedings, October, 1953.

Wechsler, D. The measurement of adult intelligence. Baltimore: Williams & Wilkins, 1944. Pp. 133, 190.

Weinberg, J. L. The workshop as a diagnostic and rehabilitative tool in vocational adjustment. Unpublished paper of Jewish Occupational Council, Pittsburgh, Pa., 1955.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

THE USE OF DEMOGRAPHIC CHARACTERISTICS IN PREDICTING LENGTH OF NEUROPSYCHIATRIC HOSPITAL STAY[1]

JAMES E. LINDEMANN, GEORGE W. FAIRWEATHER,[2] GIDEON B. STONE, ROBERT S. SMITH[3]

Veterans Administration Hospital, Perry Point, Maryland

AND IRA T. LONDON

University of Delaware

The ever-increasing number of chronic neuropsychiatric patients represents perhaps the foremost problem in the field of mental health today. Despite recent advances, notably in the fields of chemotherapy and rehabilitation, the burden of the chronic population grows, threatening to overwhelm existing hospital facilities and available professional personnel. This study was conceived as the first step in a three-part plan to: (*a*) develop an index predictive of chronicity; (*b*) attempt to isolate and study the determinants of chronicity; and (*c*) develop and test the effectiveness of retraining programs based on the defining characteristics of the chronic population.

The possibility of a predictive index was suggested by a study previously completed at the Perry Point Veterans Administration Hospital (Giedt & Schlosser, 1955). An analysis of the time sequence of discharge rate for the patient population revealed that 61% of admitted patients left the hospital during the first 90 days, 25% during the next 15 months, and only 2% during the remaining 24 months covered by the study. A predictive index, based on demographic characteristics ascertainable at the time of hospital admission, would permit a comparison of the characteristics of the "quick discharge" population with the potential chronic population. In addition, it would facilitate the early, comprehensive study of the potential chronic population so that the characteristics of those who finally did leave the hospital could be compared with those who remained indefinitely. Such an early study, before the leveling effect of hospitalization on social characteristics and attitudes became operative, was deemed particularly desirable.

The extensive literature (see, for example, Zubin, in press) on the prognosis of length of hospitalization was reviewed and the following variables selected for study:

1. Number of previous hospitalizations
2. Age at first neuropsychiatric hospitalization
3. Age at time of current hospitalization
4. Service-connection of disability (per cent)
5. Length of military service
6. Age entering service
7. Months service prior to first NP hospitalization
8. Education
9. Religion
10. Marital status
11. Number of children
12. Combat experience (Yes or No)
13. Neuropsychiatric diagnosis in service (Yes or No)
14. Diagnosis
15. Severity of external precipitating stress
16. Predisposition
17. Degree of incapacity
18. Secondary diagnosis (NP—Yes or No)
19. Legal competence (Yes or No)
20. Alcoholism (History of—Yes or No)
21. Occupational classification level

---

[1] The authors wish to acknowledge their indebtedness to John L. Holland, Albert Pepitone, and Richard Sanders for their contributions in the original formulation of this study.

[2] Now at VA Hospital, Palo Alto, California.

[3] Now at the Philadelphia State Hospital.

Method

Data for this study were derived from the clinical record folders of all male neuropsychiatric patients admitted to a large Veterans Administration hospital during the period from July 1, 1954 to December 31, 1954. With the exception of length of hospital stay, the criterion, all of the information used in the index was obtained from the initial psychiatric summary, which is completed within three weeks of admission, and from demographic data which are recorded immediately upon admission. The final total sample for the 1954 period consisted of 248 cases. A small percentage (less than 5%) of the clinical record folders were unavailable because they had been transferred to other hospitals or because they were otherwise unobtainable. Where patients had been admitted more than once during the period, data from the first admission only were included in the sample.

Information on the 21 variables that had been selected for study was recorded and the sample was divided into those who were hospitalized for 90 days or less (Short Stay group, $N = 120$) and those hospitalized for 91 days or longer (Long Stay group, $N = 128$). The high frequency of discharges found during the first 90 days in the earlier study (Giedt & Schlosser, 1955) suggested that this cutting point might have psychological, as well as statistical, significance. The two groups were then examined for differences within the 21 categories, and the significance of differences between groups evaluated by chi square.

Results

Of the 21 variables, four served to differentiate the two groups beyond the .001 level of confidence. These were: diagnosis, degree of incapacity, legal competence, and alcoholism. Marital status differentiated the two groups beyond the .01 level of confidence[4] and combat experience beyond the .05 level. A seventh variable, per cent of service connection, differentiated the groups between the .05 and .10 level of confidence. Finally, an eighth variable, occupational classification, significant beyond the .30 level, was retained in the original predictive index because of the special significance that has been accorded it (Frumkin, 1955).

A score for each item within the eight variables was computed following a method devised by Moran, Fairweather, Morton, and McGaughran (1955). This method essentially involved computing for each item the probability of a patient's falling in the Long Stay group. The log of each probability value was then ascertained. Each patient was then assigned a score which was the sum of the eight log values that applied to him.

Scores thus computed for all cases ranged from $\bar{5}.75$ to $\bar{2}.25$. Below an arbitrary cutting point of $\bar{4}.25$ were 75 cases. Of these, 85.3% were in the Short Stay group. Above a cutting point of $\bar{2}.75$ were 89 cases. Of these, 85.4% were in the Long Stay group. This index, then, served to predict long or short stay for 66.1% of the sample with a high degree of accuracy. In addition, it allowed for the specification of those cases for which accurate prediction could not be made.

Prediction for this group of 1954 admissions might have been spuriously high, of course, since the index was constructed from the very cases being predicted. Therefore, the index was cross-validated on a sample of 1955 admissions.

Cross-Validation

Data were collected for all male neuropsychiatric patients admitted to the same hospital during the period January 1, 1955 to June 30, 1955. The same scoring procedures and cutoff points were employed with this sample of 209 cases. Below the cutting point of 4.25 were 69 cases. Of these, 87.0% were in the Short Stay group. Above the cutting point of $\bar{2}.75$ were 76 cases. Of these, 78.9% were in the Long Stay group. Thus the index, based on 1954 admissions, could have predicted length of stay for 69.4% of the 1955 admissions with 82.8% accuracy.

In calculating the differential power of the individual variables on the cross-validation sample, it was found that three of the eight variables did not reach the .05 level of sig-

---

[4] Two other variables significant beyond the .01 level, external precipitating stress and predisposition, were eliminated because of their high correlation with diagnosis, which was retained in the index.

Table 1

Probabilities of Long Stay Given by Each of the Variables in the Revised Form of the Index

Variable	N^a	Chi Square p	Cutting Points	Probability of Long Stay	Log Values
Marital status	454	.001	Married and widowed	.40*	9.59988–10
			Separated and divorced	.43	9.63548–10
			Single	.66	9.81624–10
Diagnosis	447	.001	Organic	.39	9.59550–10
			Psychotic	.78*	9.89321–10
			Neurotic	.30	9.48287–10
			Character disorder	.10	9.01703–10
Degree of incapacity	445	.001	None, mild, moderate	.38	9.58320–10
			Severe	.82*	9.91169–10
			Not part of diagnosis[b]	.10	9.02119–10
Legal competence	444	.001	Competent	.33*	9.51983–10
			Incompetent	.88	9.94498–10
Alcohol	449	.001	Yes	.37	9.56585–10
			No	.65*	9.81023–10

[a] N differs for each variable because occasionally information on one item is "unknown" for a particular case. In such instances, the score used in the index is the log of the modal p value of that item, which is indicated here by an asterisk.

[b] In the case of Degree of Incapacity, "not part of diagnosis" refers to those instances, notably alcoholism and personality disorders, where this is not included in the formal diagnosis.

nificance. These were: occupational classification, service connection, and combat experience.[5] Accordingly, these variables were not included in the final, revised form of the index.

The Revised Index

The final form of the index was based on a combination of both samples, embracing all admissions during the period July 1, 1954 to June 30, 1955. Scores on the final form were derived from the remaining five variables exactly as in the original form. The individual probabilities furnished by each variable are given in Table 1.

The scores of each group distribute as shown in Fig. 1.

The validity of this index, i.e., the relationship between scores and short and long stay, clearly depends on the cutting points selected. Predictions might be made for every case by cutting above and below $\bar{2}.5$, in which case the index would yield 77.2% accuracy. Cutting scores might be established below $\bar{2}.1$ and above $\bar{2}.9$. In this case, 172 Short Stay cases would be predicted with 84.3% accuracy, and 153 Long Stay cases would be predicted with 87.6% accuracy. Thus 71.1% of all cases could be predicted with 85.8% accuracy. If homogeneity rather than size of sample is emphasized, cutting scores might be established below $\bar{3}.7$ and above $\bar{1}.1$. In this case, 102 Short Stay cases would be predicted with 91.2% accuracy, and 116 Long Stay cases would be predicted with 91.4% accuracy. Thus prediction is possible for 47.7% of the sample with 91.3% accuracy. The validity of the index might also be expressed directly in terms of the correlation between index scores and short and long stay. A biserial correlation of .759 was obtained.

Discussion

By means of this index, samples of a given size and with a given probability of long or short stay may be selected from the patients entering the hospital under study. Generalization of these results to other neuropsychiatric hospitals must, of course, await further cross-validation studies. It should be pointed out that none of the variables in the final form of the index are restricted to a veteran population, and that it has potential applicability in any neuropsychiatric hospital.

---

[5] Two of these variables, service connection and combat experience, were ones in which there was an unusually high number of "unknowns." In such cases modal scores were assigned. It is suggested that these variables may prove significant in future studies involving veterans where the data may be gathered with a higher degree of precision than was possible in this "file" study.

The value of this index for research lies in its usefulness for the economical preselection of stratified samples for intensive study during the course of hospitalization. Homogeneous populations of potential long or short stay patients may be selected for immediate study or for assignment to treatment programs for evaluative purposes.

The index is also seen to have immediate service value. In those hospitals where an intensive continued treatment program is available, those patients who fall in the Long Stay group could be transferred to such a program immediately upon completion of the routine admission procedures. Correspondingly, short-term therapeutic and counseling procedures, designed to be completed on the Admissions Service, could be instituted without delay for those patients in the Short Stay group.

In addition to the usefulness of the index for prediction, the variables themselves allow for certain speculations concerning their psychological implications. The finding that length of stay for single patients exceeds that for others suggests a possible differentiating adaptive factor. The other group, which would include the married, widowed, separated, and divorced persons, have at some time in their lives attracted and formed a relationship with another person, and thus may possess greater personal resources. An alternative explanation is that the single patients are less likely to have an established setting in the community to which to return. However, this explanation fails to account for the fact that separated and divorced patients achieved approximately the same probability of short stay as did married and widowed.

It should be pointed out that history of alcohol was scored in the study wherever it was mentioned as contributory to the current need for hospitalization, as well as where it was an established diagnosis. Several possible explanations can be advanced for the highly significant relationship between alcoholism and "short stay." In part, it may reflect a group of patients who ordinarily maintain adequate integration, but whose defenses can be temporarily weakened by alcohol, necessitating short-term hospitalization. On the other

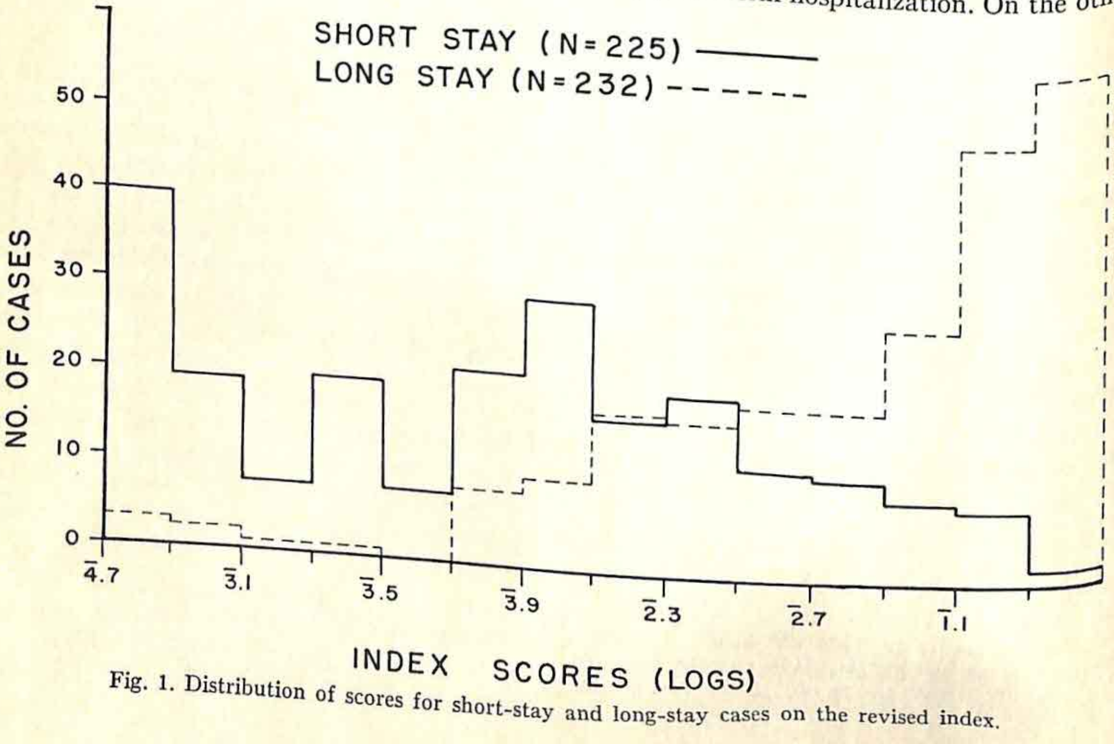


Fig. 1. Distribution of scores for short-stay and long-stay cases on the revised index.

hand, it may reflect the inappropriateness of current hospital treatment for the alcoholic. A further theoretical interpretation might be that the use of alcohol as a means of handling anxiety is generally not associated with the development of perseverant psychotic symptoms.

It is of interest that although psychiatric diagnoses have been shown to be of doubtful reliability (Mehlman, 1952), their implications for length of hospital stay have been demonstrated to be valid predictors in this study. This is seen in three of the final variables (Diagnosis, Degree of incapacity and Legal competence) which reflect the diagnostic judgments of psychiatric observers. In this case these judgments seem to reflect two underlying determinants: severity of illness of psychotics and the previously mentioned inappropriateness of treatment in the general neuropsychiatric hospital for alcoholics and character disorders. While the argument could be raised that these initial judgments merely predetermined the patients' course of hospital treatment (short or long), the accuracy of their predictive value suggests the more parsimonious explanation that the judgments genuinely reflect the patients' need for treatment in a neuropsychiatric hospital.

Summary

The potential value for research in chronic mental illness of an instrument for the early prediction of length of hospital stay was indicated. Data on 21 demographic variables, available within three weeks of hospitalization, were gathered from clinical record folders. Five variables were found to predict significantly for the initial and cross-validation samples. Based on the combined samples ($N = 457$), an index was devised which predicted Short Stay (under 90 days) and Long Stay (91 days or more). The validity of the index depends on the cutting point selected. Prediction can be made for the entire sample with 77% accuracy, or for approximately one-half the sample with 91% accuracy. Potential uses of the index and the psychological implications of the significant variables were discussed.

Received February 19, 1958.

REFERENCES

Frumkin, R. M. Occupation and major mental disorders. In A. M. Rose (Ed.), *Mental health and mental disorder.* New York: Norton, 1955. Pp. 136–160.

Giedt, F. H., & Schlosser, J. R. Movement of patients through a neuropsychiatric hospital. Unpublished manuscript, Medical Library, VA Hospital, Perry Point, Maryland, 1955.

Mehlman, B. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.,* 1952, **47**, 577–578.

Moran, L. J., Fairweather, G. W., Morton, R. B., & McGaughran, L. S. The use of demographic characteristics in predicting response to hospitalization for tuberculosis. *J. consult. Psychol.,* 1955, **19**, 65–70.

Zubin, J. Role of prognostic indicators in the evaluation of therapy. In R. W. Gerard & J. O. Cole (Eds.), *The proceedings of the conference on the evaluation of pharmacotherapy in mental illness.* (September, 1956.) Washington, D. C.: National Academy of Sciences, in press.

Bureau Ednl. Psy. Research
DAVID HA INING COLLEGE
Dated
Acss. No

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

BRIEF REPORTS

VISUAL AFTERPHENOMENA IN DIAGNOSIS [1]

DERWOOD E. JOHNSON, ROBERT W. BAUER, AND DON R. BROWN
Evansville State Hospital, Evansville, Indiana

Absence or inadequacy of spiral aftereffects has been noted in patients with organic brain involvement. There is some evidence to suggest that functional psychiatric patients also partially fail in reporting apparent motion or phi phenomena. Stationary afterimages in these groups have not been investigated.

In this research, spiral aftereffect and stationary afterimage responses were noted in normal individuals, in patients with organic brain impairment, and in patients having a functional diagnosis without indication of organic impairment, who were capable of responding to psychological examination. An attempt was made to relate response to degree of impairment or extent of morbidity by classifying "severe" and "mild" groups in both categories. A total of 100 employees and patients was used, with subgroups of 20 subjects. Age differences between groups were due to the exclusion of aged persons from the normal and functional groups. Since brain syndrome diagnoses are common among older patients, these were included in the organic sample. Nineteen records (16 of patients diagnosed as having chronic brain syndrome) were rejected because responses were doubtful.

The apparatus for the spiral aftereffect was similar to that used in previous research. To test a visual afterimage phenomenon not involving apparent motion, the "black and white" aftereffect was used. Four cards were presented: a black card with white center, a white card with black center, and gray cards for afterimages.

Group scores were evaluated for differences and Fisher's t test applied. Postdictive efficiency of the two tests was also calculated.

Differences within groups in response to the two tests were attributed to chance. Both tests yielded significantly different performance records when normal subjects were compared with groups of patients. However, neither test discriminated brain syndrome patients from functional patients at a satisfactory level, and the rate of false negatives and false positives was very high. The two tests were significantly correlated except in the severe functional group. Although the results support the conclusion that visual afterphenomena are deficient in mental hospital patients, they do not support the hypothesis that observed deficiencies are associated almost exclusively with organic brain impairment.

The low postdictive efficiency obtained in this seriously disabled population would imply no clinical utility with borderline or equivocal cases such as are commonly referred for psychological examination.

Brief Report.
Received April 11, 1958.

---

[1] An extended report of this study may be obtained without charge from Derwood E. Johnson, Evansville State Hospital, Evansville, Indiana, or for a fee from the American Documentation Institute. Order Document, No. 5779, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

INSIGHT AS A MEASURE OF ADJUSTMENT IN THREE KINDS OF GROUP EXPERIENCE[1]

JOHN H. MANN
New York University

AND CAROLA HONROTH MANN
College of the City of New York

The present study is one of a series devoted to the comparison of the relative effectiveness of group discussion, task-oriented study group activity, and group-centered role playing in producing personality and behavior change. These three group methods have been used in a wide variety of settings as a means of influencing the personalities of group members and were therefore chosen for comparison in this study.

Criteria for assessing personality change are numerous. One of the criteria recently used to measure personality change is insight. In operational terms, insight is usually defined as the degree of congruence between an individual's view of himself and the view which others have of him. Insight was used as the criterion measure in the present study. The insight scores were derived from factor analytic studies of the observable characteristics of individuals in small groups. Such studies have recently been summarized by Carter (1954), who concludes that most of the variance of small group interaction can be described in terms of the following three factors: Individual Prominence, Group Goal Facilitation, and Sociability. Two variables highly loaded on each of these factors were selected as the rating criteria on which the insight scores were based, in order to ensure the broadest coverage of independent content in the rating criteria chosen.

The study investigated the relative effectiveness of the three group methods in improving insight of group members. Ninety-six *S*s were randomly selected from a graduate course and assigned to 12 groups (3 discussion, 3 study, and 6 role playing) of eight members each. Measures of insight were obtained near the end and at the beginning of each of the group experiences which extended over a three-week period. The analysis of the data obtained from these measures indicated that (*a*) the three types of group experiences did not differ in the degree to which they increased insight of group members; (*b*) all group members showed increase in insight over the course of the group experiences; and (*c*) no relationship was found between amount of insight and individual adjustment. In addition, it was found that the role playing *S*s revised their self estimates in the direction of group estimates more often than group estimates were revised in the direction of the individual's self estimate.

Brief Report.
Received May 22, 1958.

REFERENCES

CARTER, L. F. Evaluating the performance of individuals as members of small groups. *Personn. Psychol.*, 1954, 7, 477–484.

---

[1] An extended report of this study may be obtained without charge from John H. Mann, Graduate School of Arts and Science, New York University, Washington Square, New York 3, New York, or for a fee from the American Documentation Institute. Order Document No. 5776, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 1, 1959

THE EFFECT OF EGO INVOLVEMENT AND SUCCESS EXPERIENCE ON INTELLIGENCE TEST RESULTS[1, 2]

ROBERT C. NICHOLS
Purdue University

Most discussions of intelligence testing procedures emphasize that rapport and ego involvement on the part of the *S* are important aspects of a testing situation, and that cues as to success or failure on parts of the test should not be given by the examiner. A knowledge of the probable effects of these variables is important for users of intelligence tests, since in the individual testing situation the examiner must often make allowance for lack of optimum attitudes on the part of the *S* or decide to what extent the attitude of the *S* has invalidated the test results.

To obtain data relevant to this question, 11 examiners administered the Wechsler-Bellevue Intelligence Scale to eight *S*s following a success experience, which was produced by showing the *S* false norms after taking a brief figure identification test. The other half of the *S*s were similarly given a failure experience. Half of each examiner's success and failure *S*s were given special ego-involving instructions which emphasized the importance of the intelligence test they were about to take. The other *S*s were given instructions designed to minimize ego involvement. The 88 *S*s who volunteered for the study were male undergraduate engineering students ranging in age from 18 to 32 and ranging in full scale IQ from 99 to 140. The mean IQ for the group was 124.

Considering success, ego involvement, and examiner as experimental variables, the data were analyzed according to a 2 × 2 × 11 analysis of variance design. Separate analyses were done for the verbal and performance scales, and for digit span and information plus comprehension subtests. The results are clearly negative. In the 28 *F* tests computed, no significant effects were found.

These negative findings support the assumption that differences in test taking attitude on the part of the *S* and minor differences in testing procedure on the part of the examiner do not materially affect intelligence test scores. However, since the *S*s used in this study were all intelligent students who are used to taking tests and doing their best, the results may not be directly applicable to clinic and hospital groups who may have less motivation to excel on intellectual tasks.

Since the 11 examiners used in this study had limited experience with the test, the lack of significant differences between examiners is especially significant. With normal intelligent *S*s, as used in this study, individual intelligence test scores are independent of special motivating or distracting aspects of the testing situation over a fairly wide range of conditions.

Brief Report.

Received June 30, 1958.

---

[1] The data for this study were collected by the graduate students in the writer's class in intelligence testing at Purdue University in the spring semester, 1956–1957. Those participating in the study were Irwin Monashkin, graduate assistant, Karl Beck, Paul Carlson, George Cerbus, Eugene Ebner, Gerald Engle, Judith Flannigan, Shirley Karas, Harrison McKay, Ray Mentzer, Deodandus Strümpfer, and Richard Wich.

[2] An extended report of this study may be obtained without charge from Robert C. Nichols, Department of Psychology, Purdue University, Lafayette, Indiana, or for a fee from the American Documentation Institute. Order Document No. 5777, remitting $1.25 for microfilm or $1.25 for photocopies.

Outstanding McGRAW-HILL *Books . . .*

CLINICAL PSYCHOLOGY—The Study of Persons

By Richard W. Wallen, Creelman Associates. *McGraw-Hill Series in Psychology.* 400 pages, $6.50

The major methods used by clinical psychologists are presented to help students prepare for actual practice. The treatment of the interview, projective methods, psychometric tests, and psychotherapy not only gives a survey of the clinician's work, but establishes a foundation for further, specialized courses. Case material is given to help the student become more perceptive and more able to think like professional clinicians. An instructor's manual is available for those using the text.

NONPARAMETRIC STATISTICS: For the Behavioral Sciences

By Sidney Siegel. Pennsylvania State University. 330 pages, $6.75. (Text Edition Available)

The first book-length treatment of nonparametric, or distribution-free statistics. It covers comprehensively the nonparametric statistical tests and measures of correlation, demonstrating their usefulness in behavioral science research. It is written for the reader with no special training in mathematics, and is organized to serve as a reference work and text.

SCIENTIFIC METHOD IN PSYCHOLOGY

By Clarence W. Brown and Edwin E. Ghiselli, University of California, Berkeley. *McGraw-Hill Series in Psychology.* 380 pages, $6.00.

This timely work describes the method of science as applied to the study of psychological phenomena. The text considers first the underlying logic of science as it relates to the study of human behavior, and then describes the commonly accepted steps used in conducting a scientific study of a psychological phenomenon. In the latter part of the book specific procedures and applications of the scientific methods in psychology are described.

PERSONALITY

By J. P. Guilford, University of Southern California. *McGraw-Hill Series in Psychology.* 562 pages, $7.50.

A text for upper-division psychology and education students which places emphasis on individual differences, rather than on stimulus-response psychology of individual behavior. The approach is from the standpoint of the logic of experimental method. Many of the newer methods of personality assessment are described. The author develops a model of personality based upon factor theory as a basis for the assessment and description of persons. Included are chapters describing personality in terms of dimensions found by factor analysis, including dimensions of morphology, physiology, aptitude, and pathology, as well as of temperament, needs, interests, and aptitudes.

Send for copies on approval

McGRAW-HILL BOOK CO., INC.
330 West 42nd Street, New York 36, N. Y.

ARE THERE

SOME PATIENTS

WHO DONT MAKE ANY PROGRESS?

Their basic personality structure may be incompatible with that of the therapist. Consult the new projective test.

Perceptual Personality Test (explained in the J. of Psychol., No.'s 45 and 46, 1958) by **Dr. Honkavaara** (research associate in the Department of Psychology at Brandeis University, formerly at Harvard and London Universities).

New method to reveal the basic structure of personality.

The handbook contains main principles af the new psychotherapy.

Price $10. (The profit of the test is being used to further the research).

P. O. Box 185 Belmont 78, Mass.

---

serving a *critical function* in psychology . . .

THE PSYCHOLOGICAL RECORD

a non-profit quarterly journal in theoretical and experimental psychology

EDITORIAL BOARD

PAUL SWARTZ (*Editor*), NEIL R. BARTLETT, S. HOWARD BARTLEY, J. R. KANTOR, W. N. KELLOGG, W. E. LAMBERT, HAROLD G. McCURDY, PAUL T. MOUNTJOY, N. H. PRONKO, STANLEY C. RATNER, WILLIAM STEPHENSON, EDWARD L. WALKER

Regular subscription, $4.00 per year. APA member subscription, $3.00. Student subscription, $1.50.

Mail subscriptions to:

DR. PAUL SWARTZ

Department of Psychology, University of Wichita, Wichita 14, Kansas

JOURNAL OF CONSULTING PSYCHOLOGY

June, 1959

Vol. 23, No. 3

Contents

ARTICLES

On the Breakdown of the Sense of Reality: A Study of Spatial Perception in Schizophrenia: *Ann Hozier* - - - 185

The Integrative Value of Concepts in Frustration Theory: *Norman R. F. Maier and Paul Ellen* - - - 195

Relationships Among Mothers' *n* Achievement, Independence Training Attitudes, and Handicapped Children's Performance: *Jesse E. Gordon* - - - 207

Post-Hospital Adjustment as Predicted by Psychiatric Patients and by Their Staff: *Julian J. Lasky, Gerald L. Hover, Philip A. Smith, David W. Bostian, Stanley C. Duffendack, and Charles L. Nord* - - - 213

Cues Influencing Judgment of Personality Characteristics: *Ernst G. Beier and John Stumpf* - - - 219

The Factorial Dimensions of the MMPI: *Gene G. Kassebaum, Arthur S. Couch, and Philip E. Slater* - - - 226

The Problem of Interminability in Outpatient Psychotherapy: *Donald R. Stieper and Daniel N. Wiener* - - - 237

GSRs During Repression, Suppression, and Verbalization in Psychotherapeutic Interviews: *Jesse E. Gordon, Barclay Martin, and Richard M. Lundy* - - - 243

A Note on Remoteness of TAT Figures as an Interpretive Concept: *Everett Pile, Walter Mischel, and Lewis Bernstein* - - - 252

MMPI Differences Between Parents of Disturbed and Nondisturbed Children: *Shephard Liverant* - - - 256

Responses to the Wording of Personality Test Items: *Charles Hanley* - - - 261

The Stability of TAT Fantasy and Stimulus Ambiguity: *Jerome Kagan* - - - 266 ✓

A Statistical Note on Racial Differences in the Progressive Matrices: *Arthur R. Jensen* 272

Racial Differences on Progressive Matrices: *Gerald Sperrazzo and Walter L. Wilkins* - 273

Defensiveness, Insight, and the *K* Scale: *Ewart E. Smith* - - - 275

BRIEF REPORTS

Manifest Anxiety, Intelligence, and College Grades: *Charles D. Spielberger and William G. Katzenmeyer* - - - 278

Prognostic Factors in Schizophrenia: *Beth Stone* - - - 279

Social Desirability in the Sentence Completion Test: *Vitali V. Rozynko* - - - 280

Agreement Among Raters on the Multidimensional Scale for Rating Psychiatric Patients: *C. James Klett and Julian J. Lasky* - - - 281

Validities of WAIS Short Forms in a Psychiatric Population: *William Howard* - - - 282 ✓

Psychological Test Reviews - - - 283

No. 2 wanting

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual*. Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota: [1]

[1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual*.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

ON THE BREAKDOWN OF THE SENSE OF REALITY:
A STUDY OF SPATIAL PERCEPTION IN SCHIZOPHRENIA[1]

ANN HOZIER[2]
Topeka State Hospital, Kansas

The breakdown of the sense of reality in human functioning has been viewed clinically as an essential feature of abnormal behavior. So central is the importance of the loss of reality contact in diagnosing mental illness, that the severity and the depth of this illness has been measured clinically by the severity and extensiveness of this loss of reality contact. Freud (1953c, 1953d), when he attempts to extend his psychoanalytic views to the total field of mental deviations, chooses to distinguish neurotic from psychotic behavior by this central factor of reality relationship. "Neurosis does not deny the existence of reality, it merely tries to ignore it; psychosis denies it and tries to substitute something else for it" (Freud, 1953c, p. 279).

Clearly, this loss of reality contact is not the only symptom of mental illness. Yet, because it extends to all aspects of human functioning, it is safe to say that it is of paramount importance in determining and circumscribing psychotic behavior. In this sense, then, to search for the causes of a breakdown of the sense of reality becomes practically a search for the causes of psychotic breakdowns.

For years, clinicians have been using evidences of disturbance in spatial perception as one of the manifestations of the loss of reality contact. The psychiatric examination includes such questions as "Who are you?" and "Where are you?" Confusion with regard to orientation to person and place is considered an indication of reality severance. Yet, the intimate relationship between confusion about oneself, the loss or disturbance of one's orientation in space, and the breakdown of the sense of reality has received, with few exceptions, little exploration.

Des Lauriers (1955, 1956), in an extensive study on the breakdown of the sense of reality in schizophrenic children, offers a systematic set of hypotheses concerning the essential conditions for the development of stable reality relationships. His approach touches upon the intrinsic or structural aspects of the breakdown of the sense of reality. It is less concerned with what, externally, might bring about the breakdown (family conflicts, organic brain pathology, heredity, congenital defects, etc.) than with what, internally, it involves.

The Des Lauriers hypotheses are developed in the context of the psychoanalytic theory of the ego (Federn, 1952; Freud: 1927, 1953a; Hartmann, 1957; Hartmann & Lowenstein, 1948; Rapaport, 1954). Des Lauriers hypothesizes that the primary model of reality, in the psychological experience of the individual, is the awareness of one's bodily self as bounded, finite, separated, and differentiated from everything that is not the self. The establishment of the bodily self is hypothesized by Des Lauriers to be intimately related to the amount of affective cathexis of the body.

The present study was designed to test, in part, the validity of Des Lauriers' theory, by investigating experimentally one of its logical consequences and extensions. The problem which was chosen concerned the perception of spatial relationships by the schizophrenic individual, as it is directly related to the problem of the bodily self.

---

[1] This study is part of a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy to the Graduate School of the University of Kansas. The author wishes to express gratitude to Austin M. Des Lauriers for allowing her the opportunity of exploring with him some of the implications of his theoretical ideas.

[2] Now at Mendota State Hospital, Madison, Wisconsin.

If we assume, as Des Lauriers does, that the psychological model of reality is the experience of one's bodily self as differentiated, bounded, and separated from everything that is not the self, then we may hypothesize that the loss of the sense of reality in the schizophrenic individual involves a breakdown in the bodily self as a consequence of insufficient cathexis of the body. This does not mean that the schizophrenic individual does not have affective cathexis available to him. He does, but he is not using it in cathecting his body. As long as the body is not sufficiently cathected to become bounded and differentiated from everything that is not the body, there exists no frame of reference from which to judge the reality of events in the external world or the reality of one's own psychological experiences. Consequently, as long as the self/nonself differentiation does not exist, all types of problems involved in the separation of self from nonself will be in some way affected, and among these problems is the problem of spatial relations and organization.

Methods and Procedures

Subjects

The *S*s were 25 schizophrenic women patients at Topeka State Hospital. The control group consisted of 25 women, eight of whom were employed by the hospital, two as psychiatric aides, and six as secretaries. The remaining women were housewives.

The selection criteria for the schizophrenics were: (*a*) an established psychiatric diagnosis of schizophrenia of at least two years' duration; (*b*) under the age of 60; (*c*) no evidence of neurological involvement or mental deficiency; (*d*) not under treatment with insulin coma or electric shock therapy; and (*e*) cooperative as judged by their ward psychiatrist and the investigator.

The selection criteria for the controls were: (*a*) at the present time not hospitalized in a psychiatric hospital; (*b*) under the age of 60; and (*c*) no implication of neurological involvement or mental deficiency. All of the control *S*s volunteered to participate in the experiment.

The raw score of the Vocabulary subtest of the Wechsler-Bellevue, Form I, was used as an estimate of intelligence. A comparison of the control and schizophrenic groups on the variables of age, education, and intelligence is presented in Table 1. The *F* test for homogeneity of variance with the appropriate *t* test as recommended by Edwards (1950) was used to evaluate the difference between the means of the groups. None of the means were significantly different.

The duration of illness, which was defined as the amount of time elapsed between the first diagnosed schizophrenic breakdown and the time of the study, ranged from 2 to 28 years, with a mean of 10.41 and a standard deviation of 4.15. The first schizophrenic breakdown in our population, then, occurred on the average in the early thirties. There were eight patients each of the paranoid and chronic undifferentiated type; four of the catatonic type; and one each of the hebephrenic, simple, mixed, affective, and acute undifferentiated type. The acuteness or chronicity of their symptomatology varied. Some of the patients needed maximum supervision, others minimum, and a few were in the process of obtaining positions outside of the hospital.

Table 1

Comparison of the Control and Schizophrenic Groups on Age, Education, and Intelligence

Group	Age		Education		Intelligence (Raw Vocabulary Score)	
	M	*SD*	*M*	*SD*	*M*	*SD*
Control	41.60	20.10	12.20	1.63	27.44	10.86
Schizophrenic	43.40	16.45	12.00	3.50	22.64	9.27

Experimental Materials

The Figure Placement task. The Street Scene and the Living Room Scene of the Make-A-Picture Story (MAPS) test constituted the basic models for the background pictures. These scenes were redrawn (8½ × 11 inches) in pencil so that the horizontal dimension was the main dimension of importance, i.e., the pictures were centered. The amount of structure of the scenes was changed and the size distortions corrected. An essential copy of the Street Scene is presented in Fig. 1; of the Living Room Scene in Fig. 2.

One of the male figures of the MAPS and a female figure drawn by the investigator were reproduced at three different sizes, 4⅝, 3, and 1½ inches in height. An essential reproduction of the figures is shown in Fig. 3.

The scenes and figures were sprayed with Krylon Crystal Clear Spray Coating in order to eliminate smearing. The regular stage of the MAPS with a sheet of galvanized iron in the slots, three small magnets, colored pencils, and onionskin paper constituted the remaining materials.

The Doll task. McCall's pattern No. 2000 for a 16-in. rag doll was the basic model for the construction of the dolls. The pattern was increased and decreased by one-fourth of an inch in order to make three dolls of three different sizes. The dolls had detachable legs, arms, feet, hands, and head. All parts could be joined by snaps. There were some snaps to which no body part belonged, two on the side of the head, two at the waist, and two immediately next to the head-body snap. The faces and buttons down the front of the bodies were embroidered. There was a total of 30 pieces, 10 belonging to each doll.

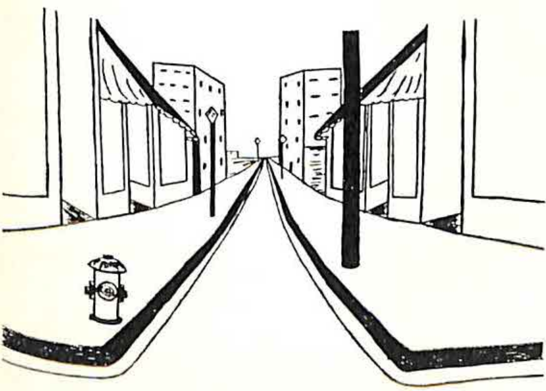


Fig. 1. The Street Scene.

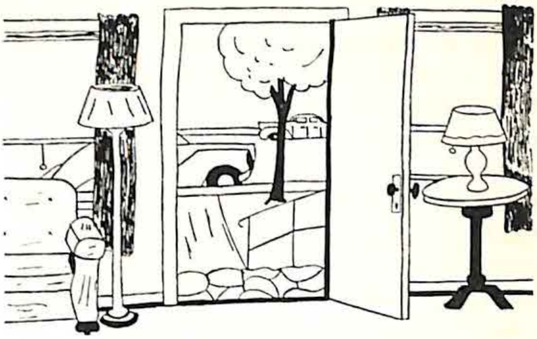


Fig. 2. The Living Room Scene.

The Draw-A-Person task (DAP). A piece of plain white paper 8½ × 11 inches and a No. *2* pencil were the materials for this task.

Experimental Procedures

All the *S*s were administered all three tasks by the same examiner in one setting lasting approximately 40–50 minutes.

The Figure Placement task. The *S*s were seated immediately in front of the MAPS' stage on which a plain piece of white paper had been placed. After a demonstration on the use of the magnets to hold the figures on

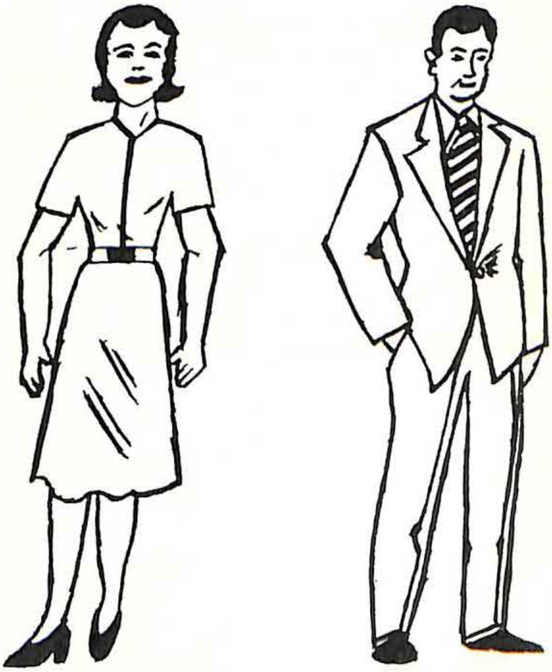


Fig. 3. The human figures.

the stage, this paper was removed and the first scene presented.

There were three trials for both scenes. On the first trial for each, all six figures were placed, in random order, on the floor of the stage. The *S*s were told to place any three of the figures on the scene wherever they looked most appropriate—wherever they looked best to them. Only the two same-sized figures were used for the subsequent trials. The Figures A ($4\frac{5}{8}$ in.), B (3 in.), and C ($1\frac{1}{2}$ in.) were presented, in that order, on both Trial 2 and Trial 3.

Each placement was recorded by placing a piece of onionskin paper over the scene and marking the lower left-hand corner of the figure with different colored pencils depending on the trial.

The Doll task. The 30 pieces of the dolls were randomly spread out. The head of the largest doll, Doll 3, was held by the examiner and the following directions given:

> There are a number of pieces here. Some of them belong with this and some of them do not. I want you to take all the pieces which belong to this and put them together with this. Just snap the pieces on. If you have any difficulty with the snaps, just let me know, and I will help you. Let me know when it looks all right to you.

The Draw-A-Person task. The *S*s were given the following instructions:

> I want you to draw a person—a whole person. Do the best you can.

Scoring of the Experimental Tasks

The Figure Placement task. The Figure Placement was scored by placing the onionskin over 10 squares to the inch graph paper. The placements were scored to the nearest tenth of an inch. The vertical dimension was recorded only to check the proper grounding of the figures.

For both scenes, Trial 1, in which three figures were placed on the picture at one time, did not figure in the final scoring. This was a practice trial to familiarize the *S*s with the backgrounds and figures and, in particular, the fact that the same-sexed figures were of the same person reproduced at different sizes —thus not a child or statue.

Three judges made judgments as to the possible range of appropriate placements for each size figure on each background. The mean of the range of their judgments was used as an index of the most appropriate placement for each figure on each background.[3]

The final score for the Figure Placement was the absolute sum of the deviations from the mean appropriate placement for each size figure on each background. There were two other errors which indicated inappropriateness of placement. Both referred to the grounding of the figures. Grounding refers to placing the figures in a stable, upright position with the feet standing on a solid surface. Figures which are tilted are standing precariously on the ground, whereas the completely nongrounded figure gives the impression of floating in the air. Only those deviations which constituted unequivocal tilts or nongroundedness were scored.

The Doll task. Connecting all the pieces belonging to Doll 3 in an appropriate manner received a score of 0. Because of the impossibility of making dolls of absolutely proportionate parts, some minor errors were expected. Omissions and deviations in placement and size were scored in terms of whether they constituted minor or major errors. What constituted minor and major errors was decided upon a priori; however, some unanticipated errors occurred, thus necessitating some addition of items. After the doll was completed, the placement and the size of the parts were indicated on a schematic diagram.

The Draw-A-Person task. The drawings were scored according to a key which was developed on the basis of a pilot study which compared the DAP of 50 college students and 50 schizophrenics, utilizing the Goodenough (1926) scoring system. The final scoring key[4] represents a modification of the Goodenough and includes those items which significantly differentiated the college *S*s from the schizophrenics, when age, educational level, and intelligence were *not* controlled. The total score for the DAP is the sum of the individual scoring points.

---

[3] The judged mean appropriate placements on the Street Scene were as follows: Figure A, 1.8 in.; Figure B, 3.6 in.; and Figure C, 4.6 in. On the Living Room Scene, the judged mean appropriate placements were as follows: Figure A, 1.6 in.; Figure B, 3.1 in.; and Figure C, 4.6 in.

[4] A 1-page scoring manual for the DAP has been deposited with the American Documentation Institute. Order Document No. 5820, remitting $1.25 for microfilm or $1.25 for photocopies.

The factors considered to be involved in space perception are position, direction, size, form, and distance (Warren, 1934). The tasks, in general, reflect varying degrees of structure in terms of spatial demands. The DAP is the least demanding. Here, the individual is confined only by the limits of the paper. He may define the human figure in any way he likes. The Doll presents more structure. It demands that the *S*s organize the human figure in relation to the head which forms the central theme of the body image (Cohn, 1953). In the DAP, he can create the form, the size, and the relation of the body parts; in the Doll, he must follow the demands placed on him by the presentation of the head of a certain size in order for the form, size, and placement of the parts to appear spatially appropriate. In the Figure Placement, the problem is much more structured. The human form is created for the *S* and he must place it in appropriate relationship to fairly well-structured backgrounds. Both the Doll and the DAP are presumed to reflect primarily the problem of spatial perception and organization in relation to the body, and Figure Placement, the spatial problem in dealing with the external world.

Hypotheses and Results

A median test with a chi square test of significance (corrected for continuity) was used to evaluate all the differences between the groups. A one-tailed test was used throughout to evaluate the level of significance.[5]

Greater variability on all tasks for the schizophrenic group is predicted because of the varying degrees of reality severance that they clinically manifest and consequently, the varying degrees of disturbance of spatial perception which is expected.

The Figure Placement Task

1. Since the schizophrenic has no intrinsic frame of reference and because he experiences himself as boundless and fluid, other objects assume this fluid, nonstable quality. When he is asked to place a well-organized human figure in appropriate relationship to a fairly well-structured background, that figure is not experienced by him as having definite form, bounds, and limits as it is experienced by nonpsychotic *S*s. Consequently, having neither the well-delimited bounds of his own body as a reference nor being able to utilize the bounds and limits of the external world defined for him, it is predicted that the schizophrenic *S*s place the figures significantly more deviant from the judged mean appropriate placement than the control *S*s.

Table 2 shows that the schizophrenics did place the figures significantly more deviant from the judged mean appropriate placement than the controls, with the exception of the placement of Figure A on the Living Room Scene. The *F* tests indicate that the schizophrenics were significantly more variable in their placement of the human figure on fairly well-structured backgrounds than the control group on the task as a whole, on placing Figure C on the Street Scene, and in placing all figures on the Living Room Scene.

2. Furthermore, since the schizophrenic is not differentiated from the world, i.e., cannot maintain adequate distance between himself and external objects, it is predicted that he place the figures more to the foreground than the controls, i.e., in the direction of fusing with them. This prediction was supported by the following data.

The distance between the average *actual* placement of Figure A and Figure C on both scenes was less for the schizophrenics than for the controls. On the Street Scene, the distance between the actual placement of Figure A and Figure C for the controls was 3.03 in., and for the schizophrenics, 2.09 in. On the Living Room Scene, the distance for the controls was 3.39 in., and for the schizophrenics, 2.79 in. That is, the schizophrenics tended to cluster the figures more than the control group. That this clustering of the figures was toward the foreground was witnessed in the less distance between the placement of Figure C and the foreground (0.00) for the schizophrenics than for the controls. On the Street Scene, the average distance between the placement of Figure C and the foreground was 4.80 in. for the controls, and 3.94 in. for

---

[5] A 1-page table giving the frequency distributions of both groups on all three tasks has been deposited with the American Documentation Institute. Order Document No. 5820, remitting $1.25 for microfilm or $1.25 for photocopies.

Table 2

Comparison of the Control and Schizophrenic Groups on the Figure Placement Task in Terms of the Deviation from the Mean Appropriate Placement

		Control ($N = 25$)		Schizophrenic ($N = 25$)						
Scene	Fig.	*M*	*SD*	*M*	*SD*	*F*	*P*	Grand Mdn.	χ^2	*P*
Street	A	1.02	1.60	1.96	1.35	1.41	—	1.54	5.36	.025
	B	1.04	1.66	2.08	2.21	1.77	—	1.00	3.62	.05
	C	0.88	1.29	1.86	3.46	7.32	.01	0.88	2.82	.05
Living Room	A	1.18	1.16	1.86	1.65	2.02	.05	1.38	0.76	.25
	B	1.04	1.18	3.40	2.80	5.64	.01	1.40	17.12	.005
	C	1.24	1.86	3.51	5.39	8.40	.01	1.83	5.88	.01
Combined Placements		6.45	4.20	15.05	12.06	8.29	.01	6.00	7.86	.005

the schizophrenics. On the Living Room Scene, the average distance was 4.69 in. for the controls, and 3.84 in. for the schizophrenic group.

3. Since the schizophrenic does not experience his body as having weight and solidity and, therefore, in need of a solid surface on which to stand, it is predicted that the schizophrenic *S*s evidence significantly more nongroundedness than the control *S*s. The relative infrequency of the two types of nongroundedness led to the decision to tabulate them together. Six of the controls and 11 of the schizophrenics gave one or more nongrounded responses, which is not statistically significant ($\chi^2 = 1.426$, $p = .15$).

The Doll Task

Body parts are not experienced as integral parts of the body by schizophrenic *S*s. The form, size, and the position of body parts are distorted. These difficulties arise because of the breakdown of the bodily self, of the experience of its size, its unity, and its extensions. Nonpsychotics, whose bodies are well-bounded, do not show such disturbances in spatial perception. Thus, it is predicted that the schizophrenics receive a significantly higher error score and that they omit significantly more body parts, show more size distortions, and more misplacements of integral parts of the body than the control *S*s.

The schizophrenics did receive a significantly higher error score ($\chi^2 = 23.16$, $p = .005$), and were significantly more variable than the controls ($F = 5.94$, $p = .01$). Table 3 shows the number of *S*s in each group giving one or more "pathological" responses on omissions, size, symmetry, reversals, and misplacements. The incidence of asymmetry was rare, whereas reversals, although they occurred, occurred with nearly equal frequency in the two groups. The minor misplacement of placing the arm on top of the shoulder rather than to the end of the shoulder, occurred infrequently, but with approximately equal occurrence in the two groups. The chi squares for the incidence of omissions of integral parts of the body, for size distortions of both a minor and major variety, and misplacements of a major type, indicated that the schizophrenics did make significantly more errors and gave significantly more "pathological" responses than the control *S*s.

The Draw-A-Person Task

It is predicted that schizophrenics show the disorganization and distortion of their bodily self and, consequently, the breakdown of spatial perception, as reflected in the drawings of the human figure, by the omission of body parts and joints, by the lack of proportion between the body parts, by transparencies, and by the lack of essential detailing which would indicate the human qualities of the figure significantly more than the control *S*s. These errors, in general, reflect the schizophrenic's disturbance in the experiencing of his body size, of his solidness and unity as a human form, and of his extensions.

Schizophrenics did score significantly lower ($\chi^2 = 8.052$, $p = .005$), and were significantly more variable ($F = 2.08$, $p = .05$) than the controls. Table 4 shows the comparison of the control and schizophrenic groups on the individual items of the DAP. Of the 24 items, 14 differentiated between the groups at the .05 level of significance or better. The presence or not of eyebrows, eyelashes, mouth, hands, arms, feet, and number of fingers did not differentiate between the groups. All of the items referring to proportion and flexibility of the body did differentiate significantly between the groups. Schizophrenics showed significantly more transparencies with reference to clothing, i.e., if clothing were present, the body could be seen through it.

Table 3

Comparison of the Control and Schizophrenic *S*s on the Doll Task

	Number of *S*s			
Items	Control ($N = 25$)	Schizophrenic ($N = 25$)	χ^2	Significance
Omissions				
1. Hands (3 ec)[a]	1	13	12.00	.005
2. Feet (3 ec)				
3. Body (6)				
4. Arms (3 ec)				
5. Legs (3 ec)				
Size				
6. One size deviant (1 ec)	12	20	4.25	.025
7. Two sizes deviant (3 ec)	2	14	11.42	.005
Symmetry				
8. Head not centered (3)	0	2	—	—
9. Arms not symmetrical (3)				
10. Legs not symmetrical (3)				
Reversals				
11. Arms reversed (1 ec)	13	14	—	—
12. Legs reversed (1 ec)				
13. Body reversed (3)				
Misplacements[b]				
14. Arm-shoulder (1 ec)	5	6	—	—
15. Arm-head (3 ec)	1	8	4.88	.025
16. Arm-waist (3 ec)				
17. Arm-thigh (3 ec)				
18. Leg-waist (3 ec)				
19. Leg-chest (3 ec)				
20. Leg-shoulder (3 ec)				
21. Leg-head (3 ec)				
22. Leg-neck (3 ec)				
23. Hand-leg (3 ec)				
24. Hand-head (3 ec)				
25. Foot-arm (3 ec)				

[a] Weighted scores for each item used in calculating total error score.
[b] The first named body part is connected with the second.

Table 4

Comparison of the Control and Schizophrenic *S*s on the Individual Items of the DAP

Item[a]	Number of *S*s: Control ($N = 25$)	Number of *S*s: Schizophrenic ($N = 25$)	χ^2	Significance
Presence of				
1. Hair	25	17	7.29	.005
2. Hair-circumference	18	6	9.70	.005
3. Forehead	25	20	3.56	.05
4. Eyebrow	19	14	1.43	.15
5. Eyelashes	6	4	—	—
6. Chin	25	17	7.29	.005
7. Mouth	25	22	—	—
8. Neck	22	17	1.86	.10
9. Shoulders	21	11	7.03	.005
10. Hands	8	6	—	—
11. Arms	25	22	—	—
12. Number of fingers	9	8	—	—
13. Feet	23	18	2.17	.10
14. Clothing	24	10	15.53	.005
Proportion				
15. Head	21	13	4.50	.025
16. Arms	21	9	10.08	.005
17. Legs	22	10	10.50	.005
Flexibility				
18. Shoulder or elbow joint	24	9	17.47	.005
19. Hip or knee joint	22	8	14.08	.005
20. Arm attachment	23	15	5.37	.025
21. Leg attachment	25	19	4.74	.025
Transparencies				
22. Clothing	20	6	13.54	.005
23. Outline of head	10	5	1.52	.15
24. None	6	3	—	—

[a] Each item contributes one point to the total score.

There was no significant difference between the groups in showing transparencies in reference to the head, i.e., seeing the outline of the head through the hair.

Discussion

The over-all results of this investigation lend support to the Des Lauriers hypotheses on the conditions which are necessary for the development of a sense of reality. The significance of the difference between the control and schizophrenic groups in terms of errors on both the Doll and the DAP was taken to indicate that the schizophrenics do have a significant problem in the perception of and organization of the body in relation to space. Despite the extreme and predicted variability of the schizophrenic group, the difficulties in spatial perception were sufficiently paramount not to be cancelled out by this variability. The results of both the Doll and the DAP indicated that the schizophrenics distort size, misplace body parts, and omit essential parts of the body.

That the schizophrenics as compared with the controls failed to indicate that the arms and legs of the human figure were capable of moving, i.e., that they were attached appropriately and had well-defined joints, is of particular interest suggesting, as we are led to believe from our knowledge of psychoanalytic ego psychology and our knowledge of the development of the bodily self, that movement is of importance for differentiating the body from the world. Freud (1953b) stresses the importance of movement in differentiating between stimuli which come from within the body and stimuli which come from without. Schilder (1953) feels that actions are of such importance in differentiating the me from the not me, that he states that if action were not possible that the body and the world would be one. In line with Des Lauriers' theoretical postulates, it is hypothesized that in schizophrenics there is a breakdown of the delineation of the body from the world. That the schizophrenics did not clearly and significantly indicate the potentialities for movement is taken as support for this theoretical postulate.

The schizophrenic is obviously capable of movement, but he tends not to experience himself as the one moving—as the prime initiator of his actions. Nor does he see the consequences of his action on another and realize that it is he who is doing thus and so. The effectiveness of Des Lauriers' therapeutic efforts with schizophrenic children in which he focusses considerable attention in pointing out the effect of movement on his part on the child, as well as the effect of movement on the child's part on him, adds support to such an understanding.

The results of the Figure Placement clearly indicated the inappropriateness of the schizophrenic's spatial perception and the severity of the problem which he experiences in attempting to relate his body to the world. Weckowicz (1957) has stated that the schizophrenics tend to overestimate distance and underestimate the size of objects seen at a distance. The results of this investigation, with the particular stimulus materials which were used, led to the opposite conclusion, namely, that the schizophrenics underestimate distance and overestimate the size of objects seen at a distance. This finding was in line with the prediction that the schizophrenics tend to fuse with the world, i.e., that they are unable to maintain sufficient distance between themselves and the world. With different stimulus materials and assuming with Ekstein and Wright (1952) some capacity for "defensive maneuvers" it may be that Weckowicz's results would be confirmed. We did not assume, in this investigation, that the schizophrenic had such a capacity for "defensive maneuvers," since it would imply that he, in the first place, was capable of appreciating distance, and only as a second thought, so to speak, removed himself to the background.

That the breakdown of the sense of reality involves a breakdown of the bodily self was witnessed by the omissions of integral parts of the body, by the disproportionate size of the body, by the misplacements and failure to maintain proper spatial relations between parts of the body, and by the inability of the schizophrenic *S*s to place the body in appropriate spatial relationship to the world. That this breakdown of the bodily self is a consequence of insufficient cathexis of the body is assumed supported by the nature of the logical deductive reasoning used in this experiment. If there were no loss of cathexis, there would be no breakdown of the bodily self; if there were no breakdown of the bodily self, there would be no disturbance in spatial relations and organization.

Certainly, the tasks are related to spatial relations and organization. What part other variables may play in accounting for the significant results cannot be determined by the present study; however, the results cannot be accounted for on the basis of age, sex, education, or intelligence since the groups were equated on these variables. Shakow (1946) has suggested that many of the significant findings which differentiate between schizophrenic and control *S*s are attributable to the lack of cooperation and interest of the schizophrenics rather than to the particular variable or variables which the experimenter is investigating. We do not think this to be true in our present research. Only those schizophrenic *S*s who were judged by their ward psychiatrist to be cooperative and who did willingly cooperate in this investigation were used. All the control *S*s volunteered and thus their cooperation was assured. We assume, then, that it is not a lack of cooperation which is determinant for our significant results.

Since it was necessary to make modifications in the Doll scoring system on the basis of the data, its sensitivity in significantly differentiating a control from a schizophrenic group must be cross-validated.

Summary

The purpose of this study was to explore one of the implications of Des Lauriers' theoretical postulates regarding the conditions necessary for the development of a stable sense of reality.

Twenty-five hospitalized schizophrenic women and 25 nonhospitalized women were administered three tasks, Figure Placement, the Doll, and the DAP, all of which involved active dealing with spatial relations and organization. Disturbances in dealing with the Doll and DAP were presumed to reflect primarily the spatial problem of dealing with the body, and Figure Placement, the spatial problem of dealing with the relationship of the body to the world.

As predicted, the schizophrenic group made significantly more errors and was significantly more variable on the over-all tasks than the controls. The prediction also was made that the schizophrenics would make significantly more errors and be significantly more variable than the control group on the component spatial factors involved in each task, such as size, position, form, and distance. For the most part, when there were frequent enough occurrences of the items, the results were significant and in the direction predicted.

The results of this investigation add to the validation of the Des Lauriers' hypotheses. Significant disturbances in spatial perception in reference to the body, and in reference to the body and the world were found. This breakdown in spatial relations and organization is interpreted as an indication of the breakdown of the bodily self. The breakdown of the bodily self is interpreted to be a consequence of a diminution of narcissistic cathexis of the body.

Received May 14, 1958.

REFERENCES

Cohn, R. Role of "body image concept" in the pattern of ipsilateral clinical extinction. *AMA Arch. Neurol. Psychiat.*, 1953, **70**, 503–509.

Des Lauriers, A. Development of the sense of reality in schizophrenic children: A contribution to the study of the structural aspects of schizophrenia. Unpublished progress report covering period from Sept., 1954 to Sept., 1955. NIMH Grant No. M-775 (C).

Des Lauriers, A. Development of the sense of reality in schizophrenic children: A contribution to the study of the structural aspects of schizophernia. Unpublished progress report covering period from Sept., 1955 to Dec., 1956. NIMH Grant No. M-775 (C).

Edwards, A. *Experimental design in psychological research.* New York: Rinehart, 1950.

Ekstein, R., & Wright, Dorothy. The space child. Paper read at American Psychiatric Association, Atlantic City, 1952.

Federn, P. *Ego psychology and the psychoses.* New York: Basic Books, 1952.

Freud, S. *The ego and the id.* London: Hogarth Press, 1927.

Freud, S. Formulations regarding the two principles of mental functioning. In *Collected Papers.* IV. London: Hogarth Press, 1953. Pp. 13–21. (a)

Freud, S. Instincts and their vicissitudes. In *Collected Papers.* IV. London: Hogarth Press, 1953. Pp. 60–83. (b)

Freud, S. The loss of reality in neurosis and psychosis. In *Collected Papers.* II. London: Hogarth Press, 1953. Pp. 277–282. (c)

Freud, S. Neurosis and psychosis. In *Collected Papers.* II. London: Hogarth Press, 1953. Pp. 250–254. (d)

Goodenough, Florence. *Measurement of intelligence by drawings.* New York: Appleton-Century, 1926.

Hartmann, H. Notes on the reality principle. In *The psychoanalytic study of the child.* XI. New York: Int. Universities Press, 1957. Pp. 31–53.

Hartmann, H., Kris E., & Lowenstein, R. Comments on the formation of psychic structure. *The psychoanalytic study of the child.* II. New York: Int. Universities Press, 1948. Pp. 11–37.

Rapaport, D. The autonomy of the ego. In R. Knight (Ed.), *Psychoanalytic psychiatry and psychology.* New York: Int. Universities Press, 1954. Pp. 248–258.

Schilder, P. *Medical psychology.* New York: Int. Universities Press, 1953.

Shakow, D. The nature of deterioration in schizophrenic conditions. *Nerv. Ment. Dis. Monogr.*, 1946, No. 70.

Warren, H. *Dictionary of psychology.* Cambridge: Houghton Mifflin, 1934.

Weckowicz, T. Notes on the perceptual world of schizophrenic patients. *Mental Hosp.*, 1957, April, 25.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

THE INTEGRATIVE VALUE OF CONCEPTS IN FRUSTRATION THEORY[1]

NORMAN R. F. MAIER
University of Michigan

AND PAUL ELLEN
Veterans Administration Hospital, Northampton, Massachuetts

In two previous articles (Maier, 1956; Maier & Ellen, 1951), the evidence on which Maier's (1949) theory of frustration was built has been critically examined in the light of alternative interpretations of the data. In both instances the evaluation of the literature revealed that the crucial aspects of the experimental data had not been satisfactorily met by other theorists. Thus one might conclude that this theory of frustration best integrates the data on which it is based, but the question of whether it does justice to other observations and experimental findings still remains.

According to this theory, an organism's behavior potentials can be described as its repertoire of behavior, which, in line with most theories, is determined by heredity, growth conditions, and experience. The expressed or observed behavior, however, is a function of both the behavior repertoire and selective processes or mechanisms. The unique aspect of Maier's theory is the postulation of two selective precesses; one operating under conditions of motivation, the other under conditions of frustration. These two mechanisms are basically different, the former selecting behavior according to the ways in which it is perceived to achieve incentives that satisfy needs; the latter, according to other principles as yet incompletely explored but different in kind from those operating under conditions of motivation. If the motivation process is operating during problem-solving performance, then frustrated behavior not only makes problem solving incidental to the needed goal object, but may actually be in conflict with it. Thus it lacks goal-orientation and therefore represents a class of behavior without a goal. Such behavior is determined largely by conditions inside the organism, so that the expression of emotions and feelings, rather than what they achieve, characteristically accompany behavior selected under conditions of frustration. The environmental condition of availability rather than need-satisfying property appears to become dominant during frustration. If behavior instigated by frustration is adaptive, at times, this is due to natural selection rather than to the organism's ability to solve a problem. Thus an angry person may strike a man who blocks his path and the man may collapse. However, if the cause of the blow is anger rather than the removal of the obstacle, the success is incidental to the purpose.

Behavior expressed under conditions of frustration is characterized by some combination of aggression, regression, and fixation (compulsion); in contrast to behavior expressed under conditions of motivation which is characterized by variability, constructiveness (maturity), and the exercise of freedom of choice. A problem situation or obstacle in the path, therefore, may either motivate an organism to show problem-solving behavior or frustrate him to show destructive, rigid, or immature behavior. The outlook in one instance is "what to do," in the other it is blaming, hostility, or self pity. Which of these types of behavior is expressed depends upon the condition of the organism, since the same objective problem may cause both types of behavior to appear in a group of individuals.

---

[1] The authors are indebted to Daniel R. Miller and L. Richard Hoffman of the University of Michigan for helpful suggestions.

The definitions of problem situations and of frustrating conditions are basically alike; the difference is in the behavior observed. This does not mean that some situations may not be more likely to frustrate people than others; rather, a crucial condition—an obstacle in the path of a goal—is a common characteristic of both. A determining difference in such situations is the amount of pressure or force exerted in requiring that the obstacle be circumvented. This pressure may serve either to motivate or to frustrate, and therefore produce opposite effects.

Although some individuals are more prone to show frustrated behavior than others, organisms are alike in that each possesses two qualitatively different behavior possibilities. This means that operational definitions of frustration and problem situations are essentially incomplete since the organism in question determines what effect obstacles and pressure will have on it. It follows, therefore, that even though situations may differ in degree, the organism's behavior will differ in kind.

The three concepts that characterize frustration theory, and therefore serve as critical areas for testing it, are (*a*) the assumption of a frustration threshold; (*b*) the postulation of a qualitative distinction between behaviors instigated by frustration and by motivation; and (*c*) the adoption of the principle of availability to account for the selection of responses that are expressed during frustration (as contrasted with the function of needs or excitatory values in motivated behavior).

These three concepts were developed in order to account for research findings obtained from the study of rats in a specific type of conflict situation. If they serve to interpret human behavior under stress, as well as animal behavior in widely different situations, there will be reason for assuming that these assumptions are fundamental principles.

Although psychoanalytic theory and frustration theory may make some of the same predictions, it is clear that the reasons supplied and the principles used will be different in origin and in kind. It is recognized, therefore that in some instances two or more theories may make similar predictions and will have to be evaluated by additional criteria. With respect to the appearance of qualitative distinctions, psychoanalytic theory and frustration theory have some common features and some differences. For example, frustration theory may predict a bimodal distribution of results, while psychoanalytic theory may assume qualitative differences (discontinuous distributions) because the concept of levels or steps is inherent in the theory. For example, a child's relations with his mother may show qualitative changes as he moves from one psychosexual stage to another. However, the concept of levels does not limit the findings to bimodal distributions in behavior data, and this difference can lead to critical tests. Another difference is that frustration theory predicts behavior differences due to two physiological mechanisms, whereas psychoanalytic theory stresses cultural (learned) factors as causes of the differences. Finally, psychoanalytic theory would make one cautious in applying the theory to subhuman forms, whereas frustration theory has its origin in the study of animals below man and stresses a biological approach. If frustration-instigated behavior has survival value it is due more to natural selection than to problem-solving or decision-making.

It is not our attempt to determine whether one theory has more universality than another or whether frustration theory is the only one that can explain certain phenomena. Rather it is hoped that the universality of its basic assumptions will be better appreciated if the theory is seen in relationship to a larger variety of data and behavior settings. Since the analysis of data, in part, depends upon the theory that is tested it is hoped that other researchers will see fit to test their findings in relation to frustration theory. At the present time the failure to test for bimodal distributions in data makes it impossible in many instances to evaluate frustration theory, and this oversight may lead to the loss of critical information.

The Frustration Threshold

Frustration theory postulates that behavior that is constructive, motivated, problem solving, or goal oriented in nature may suddenly be replaced by behavior that is hostile (aggression), immature (regression), stubborn (fixation), apathetic (resignation), or some combination of these symptoms whenever the frustration threshold of the individual is exceeded. It is also assumed that individuals will show either constructive or nonconstructive type of behavior in a stressful situation, depending on the height of their frustration thresholds.

An experiment critical of these assumptions would be one in which the introduction of something moderately threatening or stressful caused all of the members of a group to behave in the same direction. According to frustration theory, stress (punishment) can serve as a negative incentive or as a frustrating agent, depending on whether the frustration threshold of the individual has been passed. As a consequence, punishment for aggression should cause some children to show an increase in aggressive behavior, while others should become less aggressive. Furthermore, an increase in punishment should increase the amount of group aggression, primarily by increasing the number of persons whose frustration thresholds are exceeded.

There are a variety of ways in which frustration may be induced, but these do not influence the basic character of frustrated behavior. Punishment and difficult problem situations are the most commonly used experimental methods for introducing stress. The experience of failure looms in importance as one attempts to reconcile findings.

A brief examination of the literature indicates the applicability of the concept of a frustration threshold. Allport, Bruner, and Jandorf (1949) have pointed out quite conclusively that stress and/or frustration does not cause all of the members of a group to react in the same way. Analyzing the data from life histories of 90 Jews exposed to the terror of Nazism, these authors were able to identify quite distinct behavioral reactions to the stress of National Socialism in Jews in Germany. Among the reactions manifested were resignation and defeatism, regression and fantasy reactions, conformity to the new regime, aggression and displaced aggression, and planning behavior and activities aimed at coping with the threat of Nazism by getting out of Germany. Clearly here, the stress of Nazism caused behaviors in some Jews that might be considered maladjustive (resignation, regression, displaced aggression) and in others, behaviors which can be classified as constructive. The authors point out that the simple frustration-aggression hypothesis of the Yale group (Dollard, Doob, Miller, Mowrer, & Sears, 1939) is quite inadequate to account for the wide range of behaviors exhibited under stress. The variations in behavior manifested and the possibility of grouping certain behaviors as constructive and others as nonconstructive or destructive points to the need for the postulation of a threshold principle to account for apparently contradictory results.

The research of Barker, Dembo, and Lewin (1941) dealing with the effects of frustration on play behavior also reveals opposite reactions to interferences with need satisfaction. Out of 30 children, 5 showed an increase in constructiveness, whereas 25 showed a decrease in the constructiveness of the play behavior when the barrier was lowered separating the child from the highly valued toys. All of the 5 children showing an increase in constructiveness were in the group defined as the weak frustration group (those which showed relatively little time in performing frustration-instigated behavior, such as aggression against the barriers, etc.). Again, it is apparent from these data that introducing stress into a situation does not cause the same kind of behavior in all members of a group.

Other data show that the postulation of differences in frustration thresholds clarifies the appearance of splits in a population when stress is introduced. Block and Martin (1955) in a partial replication of the Barker, Dembo, Lewin experiment attempted to account for the observed differences in play behavior by relating them to some other personality variable defined as "ego-control." Ego-control was measured in terms of the ability of a child to delay gratification (collect candy without eating it), co-satiation (continue working on a task on which they had been saturated), and finally, on observed ratings of ego-control. The specific predictions made were that the overcontrolling child will exhibit less decrement in constructiveness in play between pre-frustration and frustration situations than the undercontrolling child. Hence, a negative correlation should exist between change in play constructiveness and either delay of gratification scores, or co-satiation scores. The results obtained were in the predicted direction and statistically significant.

It would appear from the consideration of these studies that some sort of threshold concept is of value in the analysis of data from studies concerned with the effects of stress and/or frustration. In addition, the Block-Martin data suggest some of the parameters that may be related to a frustration threshold. The fact that individual reactions to stress frequently are of an opposite nature indicates the difficulty of reporting results in terms of group trends. Merely reporting the average decrease in constructiveness of play as a function of frustration hides the fact that individuals reacted to the stress in different ways. The inconsistent findings on the motivating value of stress likewise point up the need for analysis of individual differences.

The postulation of a frustration threshold also may serve to explain why the Yale researchers (Dollard et al., 1939) were unable to obtain a correlation between the degree of aggravation introduced and the amount of aggression expressed. The concept of a threshold makes a change in behavior an all-or-nothing phenomenon, such as Maier and Feldman (1948) obtained when they varied the length of time rats were exposed to the insoluble problem situation.

Recently, Shimoyama (1957) repeated this study but varied the degree of frustration in the insoluble problem by using four frequencies of punishment (0%, 30%, 50%, and 70%) supplied randomly. He found that the number of rats with fixations increased with the punishment ratio, the percentages being 0, 18, 30, and 60. No differences in the strengths of the fixations were obtained, however. This means that all group differences were due to the number of animals showing the rigid behavior. Shimoyama followed the Maier-Feldman method of guidance alternated with trial and error for breaking fixations thus making it possible to measure these strengths.

The Qualitative Distinction

The assumption of a threshold concept to account for the changes in behavior that occur during frustration implies that the process underlying behavior expressed during frustration is qualitatively different from the process underlying problem-solving or goal-oriented behavior, since different behavior mechanisms would be put into operation. Thus in frustration theory it is assumed that frustration-instigated behaviors differ from goal-oriented behaviors in terms of the way in which they are related to the stimulus and behavior sequences. If two behavior mechanisms exist, behavior data should be clearly divisible into two types. The classifications *goal-oriented* and *frustration-instigated* reflect the way Maier (1949) originally divided behavior in a problem situation. Accordingly, goal-oriented behavior was described as a *means* to an *end* (goal) and which is terminated when the end is reached or the need is satisfied. Frustration-instigated behavior was described as an end to a sequence and which terminates itself when full expression is possible and further aggravation to reinstate the sequence is not supplied. Because it terminates a unit of behavior the latter type is more closely linked to the situation in which the organism finds itself than is goal-oriented behavior. According to frustration theory, the search for reinforcement, need-reducing state of affairs or purpose in frustration-instigated behavior is an unrewarding task. In fact this mental set may lead to the postulation of need systems, which often are anthropomorphic and not subject to direct empirical verification.

The value of this simplifying feature of frustration theory is very cogently illustrated by the elaborate need system developed by Ullman (1958) to account for the appearance of "compulsive" eating behavior in rats, and the one developed by Mowrer and Viek (1948) to account for eating inhibitions in an escape situation. Failure to entertain the possibility of a basic dualism at the outset forced them to assume even greater complexities and diversity later on in order to reconcile inconsistencies and to correct for theoretical discrepancies. The above illustrations taken from the experimental literature are reminiscent of the period 25 years ago when lists of instincts were developed in order to explain behavior.

The postulation of the qualitative distinction between frustration-instigated behavior and motivated behavior, together with the concept of a frustration threshold also makes it possible to put some order into the theoretical difficulties created by the data obtained from avoidance conditioning experiments. According to frustration theory, a noxious stimulus acts as a negative incentive unless or until it is of such duration or intensity that it crosses the frustration threshold of the subject. The learning and extinction of avoidance and approach responses should obey the same laws, just so long as frustration does not enter the picture. However, if our assumptions are correct, frustration is more likely to complicate the picture when avoidance behavior is investigated. This prediction does not follow from motivation and learning theories.

In the case of the extinction of avoidance behavior, the data of Solomon and Wynne (1954) are of interest. They report that after only 3 or 4 shocks which were just subtetanizing and with a 10-sec. CS-UCS interval, the instrumental avoidance response of jumping a barrier became more and more stereotyped and the latency of the response became shorter (levelling off at 1.6 sec.). Dogs typically continued to respond to the CS for several hundred trials without showing any signs of extinction. Additional data by Solomon, Kamin, and Wynne (1953), however, show that under certain conditions extinction of these responses can occur, although the extinction was not complete in that some of the animals still failed to extinguish. Thus, when a combined procedure of preventing the response by placing a glass barrier in front of the hurdle and punishing the response was used, 14 out of 16 animals were able to extinguish their conditioned avoidance responses. Certain additional features of the experiment remain to be pointed out. First, all these animals had an intervening extinction procedure before the combined glass barrier and punishment extinction procedure was used. In six of the animals, a glass barrier extinction procedure alone was used for 10 days prior to the introduction of the combined extinction procedure, and for 10 of the animals a punishment extinction procedure alone was used for 10 days before the combined extinction procedure was instituted. When the data are analyzed in terms of this procedural breakdown, it was found that all 6 of the animals which received the glass barrier procedure prior to the combined extinction procedure extinguished their avoidance responses, whereas only 8 out of 10 of the animals which received the punishment procedure extinguished. Furthermore, the mean number of punishment shocks in the combined extinction procedure reflected the influence of whether the prior extinction procedure was the glass barrier or the punishment. That is, the six dogs that had the glass barrier prior to the combined procedure required only 8.2 shocks in the combined procedure, while the animals that had the punishment procedure required an average of 14.3 shocks, a difference which is significant at the .01 level. The important facts of this investigation that require explanation are (*a*) the failure of ordinary extinction procedures (omission of the UCS) to cause extinction, (*b*) the differences in extinction between animals receiving the glass barrier treatment and animals receiving a punishment extinction procedure prior to the combined procedure, and, finally, (*c*) the failure of extinction in some cases even under the best of the extinction procedures. Solomon and Wynne's principle of the conservation of anxiety seems adequate to account for the first fact—the failure of extinction to occur when merely the UCS is omitted. In other words, the short-latency avoidance reactions prevent the CS from arousing anxiety reactions, thereby conserving conditioned anxiety reactions from extinction. Also it can handle instances in which extinction occurs when only the glass barrier extinction procedure is used. However, the principle does not account for the fact that some animals never extinguish even under the best of extinction procedures. This fact of the bimodality—the split in the population—is directly explained by frustration theory because it postulates the existence of a frustration threshold. Thus animals that become frustrated by the UCS will never extinguish in the usual way, whereas those with thresholds sufficiently high to resist the frustrating effects of the UCS will be able to extinguish under certain experimental conditions. Furthermore, the difference in the proportion of animals extinguishing among those exposed to the glass barrier procedure and those exposed to the punishment procedure prior to the combined procedure also can be readily explained in terms of the dual effects of punishment postulated by frustration theory. Thus, the animals exposed to the punishment procedure prior to the combined procedure received more punishment for the avoidance response than did animals exposed to the prior glass barrier procedure. According to frustration theory, this additional punishment served to make the former group rigid (i.e., resistant to extinction), because a greater number of the animals had their frustration threshold exceeded. The increased resistance manifested itself both in the number of shocks required in the combined procedure and in the smaller number of animals extinguishing.

Frustration theory does not require that all rigid behavior have its basis in frustration. Rather, this condition is but one of several under which the behavior of an organism may exhibit persistence. Other conditions are overlearning, the operation of special needs, perceptual differences, and low intelligence. Thus, the failure of extinction in avoidance behavior may either reflect the operation of a frustrated condition, and thus be an instance of abnormal fixation, or simply result from the lack of a perception of more adaptive alternatives, due to the nature of the particular experimental situation. A learning-motivation concept, such as the conservation of anxiety, readily accounts for the latter alternative but is inadequate to account for the first.

The dual operation of noxious stimuli (either as a negative incentive or as a frustrating agent) should also be reflected in other types of learning experiments that utilize them as training aids. If parameters such as intensity of noxious stimulus or the number of reinforced repetitions are used, learning and motivation principles would predict a whole spectrum of response strengths when a random population is tested. However, frustration theory would predict a bimodal distribution of response strengths; fixated and nonfixated. For example, if shock were used as a driving stimulus, it would be predicted that the percentages of animals falling into the fixated group within a given level of shock intensity would increase as the shock intensity increased, but the strength of the fixation would not change. If such an experiment yielded several groupings of response strengths that were a function of the intensity of the noxious stimulus, a case could be made against the necessity for postulating a frustration threshold and the qualitative difference between fixated behavior and ordinary habits.

The part played by needs in determining response strength might be tested by using different kinds of noxious stimuli. If the results yielded as many groupings of extinction scores as types of noxious stimuli used, then a strong case would be made for the viewpoint that response strength is a function of the type of motivation used and that differences in types of motivation can cause splits in a population. Frustration theory, however, would demand bimodal distributions within each type of noxious stimulus.

A study by Hammes (1956) may be used to illustrate the dual effects of noxious stimulation when used as a training device. In an experiment utilizing two levels of task difficulty and three levels of shock motivation, rats had to learn to escape from a shock box through the one of two doors that had the correct card on it. Performance was measured in terms of two criteria: (*a*) the number of trials to learn and (*b*) the number of errors made. In the easy task (black-white discrimination) three intensities of shock motivation (low, medium, and high) caused only one failure to learn (i.e., reach a 10 successive errorless trial criterion) in each condition. However, in the difficult task (discrimination between two circles with a 4–1 size difference) the number of animals failing to reach the criterion of learning increased from 6 to 10, as the shock intensity increased. Of interest from the point of view of frustration theory is the fact that (*a*) bimodal distributions were obtained with each of the different levels of motivation used, and (*b*) the number of animals falling into the failure group increased as the level of shock was raised. Hammes, however, treated his data in terms of group averages, and as a result these important aspects of the data were not emphasized in his paper.

Behaviors that are similar on the surface but that have different origins also may be expected to show basic differences when further analyzed. For example, crimes may be committed because they are perceived as the solution to a problem for the individual or because they reflect a hateful or antisocial condition. It is possible, therefore, to consider some crimes as either motivational or the product of frustration. If this is the case, cure and prevention would differ depending on the type of origin. It is also possible that these two processes may supplement each other since the behavior they elicit is not necessarily in conflict.

A paper by Jenkins (1955) describing two qualitatively different types of delinquency supports this analysis. In his treatment he differentiates delinquency as an adaptation from delinquency as a maladaptation. In the former instance, delinquent behavior differs from normal behavior only in its illegality. Its motivational aspects (acquisition, planning, integrated action, etc.) are the very ones which our culture sanctions and which keep our competitive economic system functioning. However, the delinquent behavior which he categorizes as maladaptive is "an explosion of blind destructiveness in which the individual utterly loses control of himself" (p. 10). It represents a renunciation of goals rather than the selection and pursuit of a goal. Jenkins further asserts that these two types of delinquency have a different origin. In adaptive delinquency there is evidence of a lack of paternal guidance but with a maternal care adequate for developing social responsiveness. In the maladaptive delinquency, however, there is evidence of primary need frustration. Typically the child is unwanted by both parents, particularly the mother, with the result that no social responsiveness can develop. Finally Jenkins points out that the distinction has definite implications for treatment. In the case of adaptive delinquency, the emphasis must be placed upon thwarting the delinquent activity. In the maladaptive delinquent, the emphasis must be placed on reducing the frustrating conditions that create the delinquency. As an example, he emphasizes the development of warm interpersonal relationships, which must be maintained despite unacceptable behavior.

Observations of cases of problem readers likewise support this approach. It is well known that therapy helps some problem readers more than drill in reading, whereas for other children the reverse is true. Research is needed to test and evaluate these observations. If frustration theory is to serve as a guide it follows that mentally retarded children need motivation and drill; whereas frustrated children need therapy to get rid of their frustrations, including the resultant regression.

The Principle of Availability

Frustration theory postulates that when an organism is in a state of frustration, the mechanism for selecting or determining behavior differs from the one that operates during conditions of motivation. The theory states that the specific character of the response elicited will be determined by its degree of availability to the subject, and that the response will have the property of aggression, regression, fixation-rigidity, or some combination of these.

Although the mechanisms for response selection under conditions of motivation vary somewhat with the point of view of the theorist, the factors usually mentioned include excitatory value of stimulus, the needs of the individual, the previous training of the individual, and certain perceptual factors. Although availability may be implied in some of these instances it is not usually regarded as a determining influence largely imposed by the environment. Rather, response selection under motivation is consistent with the psychology of choice behavior, which in turn gives the organism a determining role in its own behavior. Behavior elicited during frustration, in contrast, takes on a character of compulsiveness so that the organism may behave in a manner that is inconsistent with its choice.

Factors which make one response more available than another require further investigations to clarify the limitations. However, enough is known to include conditions such as physical nearness, biological ease (including the response in progress), natural or unlearned preferences, primitiveness or simplicity of response, previous experience or training, and cultural or sociological influences.

Although these factors are basically different, they have one thing in common: they make certain responses more accessible to the organism than others. However, the fact that there is a multitude of determinants makes complete prediction difficult. One can test them only by equalizing the other determinants while testing one. Whether it will eventually be possible to assign weightings to the factors so as to increase prediction remains a question. It is likely that individual differences will exist and that the weightings will fail to generalize completely.

Physical Nearness

According to this principle, things or persons near to the frustrated individual will tend to be involved in the behavior to a greater degree than other objects. Thus a frustrated person is likely to vent his hostilities on a person physically close to him. Likewise the frustrated person is inclined to use objects that are handiest for attack, rather than most efficient. For example, a woman may throw a plate while drying dishes but will strike out with a broom while sweeping.

This factor is consistent with the observation that a frustrated individual often attacks innocent bystanders, which is a puzzle from the view-point of problem-solving. It also explains why the source of frustration sometimes is attacked since, in many instances, the object of frustration is near to the person who is frustrated. Banging on locked doors during fires or while imprisoned, even to the point of causing self-injury, is not uncommon.

Specific experiments testing this factor have not been designed, but there is a good deal of observational data to support it. Barker, Dembo, and Lewin (1941) described children kicking teachers and barriers, and the Yale group (Dollard et al., 1939) described how subjects by their drawings vented their frustration on the experimenters. The reported negative correlation between the price of cotton and the number of negroes lynched in the South is also in line with this factor.

Vandalism and thefts of useless objects, which are difficult to explain on a problem-solving or motivational basis, may readily lend themselves to explanation in terms of this factor.

Biological Ease and the Response in Progress

Other things being equal, the response selected under frustration is likely to be one that is convenient for the individual because of his make-up or particular environmental circumstances at the time of frustration. Pushing is easier than pulling, and kicking is easier under some conditions than others. Thus the form of aggression may vary with individuals and with circumstances. Response fixation may also be influenced by ease or convenience. A forward movement is easier than a turn, and continuation of a response is easier than a change. For this reason, the response in progress at the time of frustration tends to be highly accessible, unless the conditions of frustration block it in some way.

An experimental demonstration of this factor was made by Kleemeier (1954). He trained some rats to go right, others left, and still others to go straight ahead at a junction in a maze. Mastery of one of these responses was their first habit. He then tried to develop a second habit by punishing the rats with electric shock for following the first habit. It was found that punishment fixated responses but that other factors determined which response became fixated. Usually either the old response (the one in progress) persisted or a response to the alley lying straight ahead (the one most accessible) was fixated. When the forward alley was the initial response, it was seldom abandoned. This experiment nicely demonstrates how the physical arrangement of a situation and the response in progress makes certain behaviors psychologically more accessible to the subjects.

Another example is the "compulsive" eating symptom described by Ullman (1951). His experiment illustrates the fixation of an on-going response, and one need not postulate the highly complex motivational scheme such as Ullman develops in order to account for its occurrence. The literature on perseveration (Everall, 1935; Hamilton, 1916; Hamilton & Krechevsky, 1933) and on rigidity is rich in data supporting the principle that the response in progress at the time of frustration is highly susceptible to fixation.

Unlearned Preferences

Personal or species preferences, even when unlearned, constitute another factor related to availability. Animals may have preferences for brightness, color, position, etc. when presented with discrimination choices, and it may be assumed that such preferences will influence the response selected during frustration.

An experiment by Aebli (1952) supports this factor. Using a large cross maze, unlearned preferences for left- or right-hand turns were determined in 11 rats. With the aid of a differential delay (60 sec. vs. 5 sec.), 6 animals were trained in the direction opposite their initial preference, while 5 were trained in the direction of their initial preference. After this training, shock was administered at the choice point. Five of the six animals trained on the opposite side of their initial preference regressed to their initial preference, whereas four out of the five trained in accordance with their initial preference persisted in that response. When a relearning attempt was made with the animals that regressed to their initial preference it was found that these animals were fixated and did not relearn a response that was opposite to their natural preference. This experiment suggests that prior preferences in animals act as strong determiners of behavior under stress conditions and tend to make certain behaviors more accessible than others.

Simplicity of Response

A simple response is more likely to become fixated, other things being equal, than a complex or involved one. In the Michigan experiments (Maier, 1949), it was the consistent finding that rats fixated position responses more often than discrimination responses. According to this factor it follows that an act of aggression is more likely to be a gross or primitive type of movement than a fine skill. As a corollary to this hypothesis it would be expected that complex activities should be more susceptible to stress variables than simple activities; and they should reflect this increased susceptibility both in terms of differences in rates of acquisition and in a more rapid deterioration under stress.

With this frame of reference, the material on the influence of drive level on the acquisition of simple and complex activities can be considered in a different manner, and some of the contradictions in the data can be resolved. It will be recalled that Taylor (1951) had postulated that anxious *S*s operate at a higher drive level than nonanxious *S*s, and since response strength is a multiplicative function of habit strength and drive, high anxious *S*s should learn more rapidly than low anxious *S*s. This was confirmed for a simple task, such as a conditioned eyelid response. However, Montague (1953) showed that anxiety actually interfered with verbal learning (a more complex task). As a result of this finding Taylor and Spence (1952) modified the earlier theory in order to take into account the complexity of the task and postulated that higher drive states increase the number of competing responses that will interfere with the correct response in a complex problem. More recently though, Saltz and Hoehn (1957) demonstrated that when materials in which there are competing and noncompeting response tendencies are equated for difficulty, the Taylor-Spence prediction was not substantiated. Rather, the data pointed to the likelihood that task difficulty rather than the operation of competing response tendencies was the more important variable. This finding of Saltz and Hoehn is entirely in accord with the expectations of frustration theory that make a complex or difficult task more susceptible to the influence of stress than a simple task. Thus, there is no necessity for postulating complex response interference processes, such as Taylor and Spence (1952) had done, to account for the differential effects of anxiety (if one treats anxiety as a stress variable) on simple and complex learning.

Finally, the previously cited paper by Hammes (1956) also may provide evidence for this point of view. On the easy task, varying shock intensity causes no significant differences on the acquisition of the response, whereas on the difficult task high shock required significantly more trials for learning, and significantly more errors occurred than for the low shock conditions.

Training

Kleemeier (1942) found that a frustrated rat will fixate a previously learned alley in preference to a new one if other things are equal. However, this tendency is not as strong as is usually supposed, and is off-set by the greater importance of the response in progress and the ease of movement. Nevertheless, it must be included as a factor making for response availability during frustration. According to this principle, a trained boxer is more likely to hit with a skilled jab than with a haymaker when engaged in a barroom brawl, while a wrestler in the same situation would be more inclined to select a haymaker.

This factor perhaps would not be crucial for testing frustration theory because learning is regarded as an important selector of behavior in most theories. Frustration theory makes learning a less important factor, however, and in this respect it might become an issue for a critical experiment.

Sociological and Cultural Influences

Although cultural influences are learned, the effect of culture is somewhat more complex than that of learning. One important complication is that social pressure operates to train and to motivate a social individual to follow the behavior pattern of his group. Although research in this important area is sketchy and inconclusive it should be explored because of its potential theoretical importance.

According to this factor various cultural groups should express their frustrations differently. When frustration occurs, males should be more likely to show aggression than females in certain cultures but not in others—depending upon the special values of the culture. It also follows that the proportion of physical fighting to verbal attacks in playgrounds should vary in different social groups, being greatest in groups where a given form of aggression is most acceptable. The objects selected for attack as well as the mode of attack likewise will be influenced by cultural factors. Thus scapegoats for attack may be Negroes, Mexicans, Japanese, Irish, Jews, etc. depending on locality or cultural group in which the frustration occurs. This cultural influence would supplement the factor of nearness.

The research of Miller and Swanson (in press) supplies many illustrations of the manner in which middle- and working-class groups respond to stress. Because of the cultural differences between these two classes they reasoned that working-class males should be more inclined to express themselves by using their skeletal muscles, while middle-class males should be more inclined to utilize conceptual modes. Elton McNeil, a member of the project, obtained experimental verification for this hypothesis, using matched groups of boys from each of the two classes as subjects. In a series of experiments reported in this volume many class differences were obtained, the behaviors tested ranging from levels of aspiration to expressions of aggression.

One of the differences, which was primarily a function of class rather than child-rearing practices, was found when aggression and conscience were in conflict. Middle-class boys develop defenses that affect the control of aggression if made anxious about aggressive tendencies. These defenses take the form of reversal of feeling, turning on the self, and distortions regarding the attraction of forbidden acts. Working-class boys, however, are not inclined to develop defenses against aggression and seem less concerned with its control.

When no anxiety against aggression is introduced into the situation, working-class boys express their anger in the form of physical attack, whereas middle-class boys tend to engage in verbal argument or some form of conceptual aggression. In this instance, child-rearing practices are relevant. Working-class mothers use corporal punishment to deter their sons from fighting, who in turn use and advocate direct aggression. They learn the opposite of what their training is intended to accomplish. Middle-class mothers use psychological discipline, and their sons in turn follow and advocate a similar pattern of discipline.

Miller and Swanson regard the pattern of home discipline to be one of the distinguishing features between these two classes, and they believe that learning plays a predominant influence in determining the obtained differences. However, it can also be concluded that this early training establishes values and attitudes which, in turn, may inhibit the expression of certain behavioral modes and consequently make others more available to the child. It is clear that these early learnings are not specific in that the child is taught to behave in a certain manner; rather he acquires attitudes and values from his background and these select the behavior.

The Miller-Swanson studies include experimental data based on projective tests as well as observational data. For the further development of frustration theory it would be desirable to separate expressions of true frustrations from those revealed in test situations.

Ecological, class, and cultural differences in the incidence of various kinds of behavior pathology also may be explained by this cultural factor. Rennie and Stole (1956) find that various kinds of psychosomatic disorders, such as asthma, hay fever, arthritis, heart conditions, high blood pressure, etc., differ considerably in their distribution along a socioeconomic status range. Thus, hypertension, arthritis, neuralgia-sciatica increase in prevalence with progressively lower socioeconomic status, whereas colitis, hives-rash, and hay fever increase in prevalence with higher socioeconomic status. Apparently there are factors in different socioeconomic levels that predispose individuals to particular psychosomatic symptoms. That this predisposing factor is not a motivational one is suggested by the fact that no relation between tension-anxiety and socioeconomic level was obtained. In a similar vein, Odegaard (1956) reports that various kinds of occupations predispose individuals to particular behavior pathologies. The general conclusion therefore may be drawn that the symptomatology displayed by emotionally disturbed peoples varies with class, sex, and socioeconomic status.

Sex, Body Structure, and Personality

Sex differences and variations in body build also may be expected to influence the availability of responses. It is well known that an unhappy home life causes boys and girls to express their frustrations differently. Although part of these differences may be due to training or cultural influences, it is best to keep this category separate from others until such related factors are found to be fully overlapping.

One relevant experiment related to this factor is that of Marquart (1948). She observed men to be more likely to show hostility (aggression) in her frustration experiments than women, while women were more inclined to cry (regression) than men.

The influence of body structure could best be investigated by researchers with different animal forms. It will readily be conceded that cats, deer, rats, chickens, pigs, etc. respond differently to punishment, capture, and attack, but no comparative studies have been made to formulate principles. It also may be supposed that a man's size, strength, and physical handicaps will influence his behavior during frustration, even after his training and culture influences have been discounted.

The suggestion by Jenkins (1955) concerning the influence of the kind of musculature and degree of physical vigor on the nature of maladaptive delinquent behavior is relevant here. In those children who have good musculature and a high level of physical vigor, directed aggressive attacks may be the child's way of contacting the world. In those weaker and less well equipped for such behavior, a "screaming loss of control" with a hostile resentful overdependence may result. In this instance the prognosis is towards psychotic disorganization, whereas in the aggressive type the end result is usually the aggressive psychopath.

Although culture and training overlap with conceptions of personality, it should be listed as a separate factor. Personality may determine whether the first symptom of frustration will be aggression or regression. Marquart's (1948) study indicated that extroverts were inclined to show aggression, whereas socially dependent individuals were inclined to show regression when frustrated. More research is needed to clarify this factor.

Summary

The present paper has dealt with the three major explanatory concepts of frustration theory: the frustration threshold, the qualitative distinction between frustration-instigated behavior and motivated behavior, and the principle of availability. Illustrative data from the literature of various research areas were analyzed from the point of view of the theory to demonstrate how its concepts could be utilized in explaining and integrating these diverse kinds of data. In addition, an attempt was made to determine whether these concepts could clarify research areas that were confused by contradictory data. This was done by attempting to integrate the contradictory data within the framework of the theory and also by suggesting questions that might be subject to empirical verification.

The factors that determine the form of expression that frustration will take and the types of behavior pathology that may appear were examined. It is clear that the nature of the provoking conditions, the types of needs that are denied, and the intensity of need deprivation are insufficient variables to account for the findings. Research in this area is badly needed and it is hoped that the issues raised will suggest additional problems to explore and new categories of data to tabulate.

Although the present elaboration of the concepts of frustration theory led to an explanation and integration of material from a variety of sources, there is no intent to convey the impression that frustration theory is a completed system. Many details and relations need to be extended and clarified. Since one of the requirements of a meaningful behavior theory is that it be helpful in generating testable hypotheses and concepts, it is hoped that in pointing up crucial issues in a wide range of data, differences in theoretical positions can be subjected to experimental investigation and hence to a further elaboration of the theory. A prematurely closed theory may be a handicap to new research because it delimits the regions explored. What is needed is the location of new areas.

Received April 1, 1958.

REFERENCES

Aebli, H. Regression towards an unlearned response. *J. comp. physiol. Psychol.*, 1952, **45**, 258–263.

Allport, G. W., Bruner, J. S., & Jandorf, E. M. Personality and social catastrophe. In C. Kluckhohn & H. A. Murray, *Personality in nature, society and culture*. New York: Knopf, 1949.

Barker, R., Dembo, T., & Lewin, K. *Frustration and regression: An experiment with young children*. Univer. Iowa Press, 1941.

Block, Jeanne, & Martin B. Predicting the behavior of children under frustration. *J. abnorm. soc. Psychol.*, 1955, **51**, 281–285.

Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., & Sears, R. R. *Frustration and aggression*. Yale Univer. Press, 1939.

Everall, E. E. Perseveration in the rat. *J. comp. Psychol.*, 1935, **19**, 343–369.

Hamilton, G. V. A study of perseverance in primates and rodents. *Behav. Monogr.*, 1916, 3. P. 65.

Hamilton, J. A., & Krechevsky, I. Studies in the effect of shock upon behavior plasticity in the rat. *J. comp. Psychol.*, 1933, **16**, 237–253.

Hammes, J. A. Visual discrimination learning as a function of shock-fear and task-difficulty. *J. comp. physiol. Psychol.*, 1956, **49**, 481–484.

Jenkins, R. L. Adaptive and maladaptive delinquency. *Nerv. Child*, 1955, **11**, 9–11.

Kleemeier, R. W. Fixation and regression in the rat. *Psychol. Monogr.*, 1942, **54**, No. 4.

Maier, N. R. F. *Frustration: The study of behavior without a goal*. New York: McGraw-Hill, 1949.

Maier, N. R. F. Frustration theory: Restatement and extension. *Psychol. Rev.*, 1956, **63**, 370–388.

Maier, N. R. F., & Ellen, P. Can the anxiety-reduction theory explain abnormal fixations? *Psychol. Rev.*, 1951, **58**, 435–445.

Maier, N. R. F., & Feldman, R. S. Studies of abnormal behavior in the rat: XXII. Strength of fixation and duration of frustration. *J. comp. physiol. Psychol.*, 1948, **41**, 348–363.

Marquart, D. I. The pattern of punishment and its relation to abnormal fixation in adult human subjects. *J. gen. Psychol.*, 1948, **39**, 107–144.

Miller, D. R., & Swanson, G. E. *Inner conflict and defense*, in press.

Montague, E. K. The role of anxiety in serial rote learning. *J. exp. Psychol.*, 1953, **45**, 91–96.

Mowrer, O. H., & Viek, P. An experimental analogue of fear from a sense of helplessness. *J. abnorm. soc. Psychol.*, 1948, **43**, 193–200.

Odegaard, O. The incidence of psychoses in various occupations. *Int. J. soc. Psychiat.*, 1956, **2**, 85–104.

Rennie, T. A. C., & Stole, L. Social class prevalence and distribution of psychosomatic conditions in an urban population. *Psychosom. Med.*, 1956, **18**, 449–456.

Saltz, E., & Hoehn, A. J. A test of the Taylor-Spence theory of anxiety. *J. abnorm. soc. Psychol.*, 1957, **54**, 114–117.

Shimoyama, Takeshi. Studies of abnormal fixation in the rat: I. Effects of frequency of punishment in the insoluble situation. *Jap. J. Psychol.*, 1957, **28**, 203–209.

Solomon, R. L., Kamin, L. J., & Wynne, L. C. Traumatic avoidance learning: The outcome of several extinction procedures with dogs. *J. abnorm. soc. Psychol.*, 1953, **48**, 291–302.

Solomon, R. L., & Wynne, L. C. Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychol. Rev.*, 1954, **61**, 353–385.

Taylor, Janet A. The relationship of anxiety to the conditional eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

Taylor, Janet A., & Spence, K. W. The relationship of anxiety level to performance in social learning. *J. exp. Psychol.*, 1952, **42**, 61–64.

Ullman, A. D. The experimental production and analysis of a compulsive eating symptom in rats. *J. comp. physiol. Psychol.*, 1951, **44**, 575.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

RELATIONSHIPS AMONG MOTHERS' *n* ACHIEVEMENT, INDEPENDENCE TRAINING ATTITUDES, AND HANDICAPPED CHILDREN'S PERFORMANCE

JESSE E. GORDON
Montana State University

The research reported here was designed to test three related hypotheses concerning the variables influencing the performance of hearing handicapped children.

The particular performance in which we were interested may be described as the discrepancy between a child's intellectual ability, as measured by the Merrill-Palmer, and the extent to which he has accomplished certain developmental tasks which contribute to his independence from home and parents, as measured by the Vineland Social Maturity Scale. Rehabilitation workers are interested in the variables which influence the extent to which handicapped children make maximal use of their resources. The Merrill-Palmer—Vineland discrepancy (MP-V) seems to be a reasonable measure of this ability-performance relationship.

The three sets of variables whose relationships to MP-V were studied were: (*a*) maternal independence training attitudes toward the deaf child; (*b*) selectivity in such attitudes on the basis of the hearing handicap; and (*c*) mothers' achievement needs.

McClelland, Atkinson, Clark, and Lowell (1953) have presented evidence from Winterbottom (1953) demonstrating that achievement motivation in boys is related to the independence training attitudes of their mothers. It appears reasonable that such maternal attitudes might also be related to MP-V performance in handicapped children, on the assumption that MP-V reflects, to some extent, the children's drive to achieve, as well as their tendencies to become independent. Although Winterbottom studied the relationship between independence training attitudes and *n* Achievement only in boys, adequate evidence for sex differences in this relationship does not exist. While mean level of independence training attitudes may differ on the basis of the sex of the child, and MP-V may also show sex differences (although Doll did not find them in his standardization of the Vineland), there is no reason to suppose that the relationship between the variables should differ as a function of sex, particularly if independence training attitudes represent an antecedent condition for the behavior being investigated. Therefore the present research uses deaf children regardless of sex.

A major problem in Winterbottom's study is the retrospective nature of the mothers' responses to the independence training questionnaire. One might well argue that, rather than early independence training being a causative element in the production of *n* Achievement, the situation might be reversed. That is, the mothers of high *n* Achievers may have responded to the questionnaire by reference to the ages at which their children accomplished the developmental tasks which constitute the items. It may be that high *n* Achievement children spontaneously "demanded" independence early, and in this way influenced the mothers' retrospective age estimates. The research reported here tends to minimize this problem through the use of children who are too young to have accomplished most of the developmental tasks listed in the questionnaire. Mothers' responses are therefore less influenced by past experience with their children. To the extent that this is so, a purer measure of *attitude* is obtained.

The second hypothesis studied here was the favoring later independence, when the groups are matched in terms of number of male and female children (Table 2, Line 1). However, a plot of the data very clearly indicates a curvilinear relationship between these two variables. All 19 mothers were ranked on their independence training attitudes toward normal children with Rank 1 assigned to the mother with the earliest mean age on the independence training questionnaire, and then on *n* Achievement, with Rank 1 assigned to the mother with the lowest score. When the ranks on the two variables are compared, it is found that among the mothers ranked from 6 to 15 on independence training are found all but one who ranked below 10 on *n* Achievement. There is no appropriate statistic to describe this relationship where the *N* is so small. However, the data do suggest that mothers with low achievement motivation tend to be more moderate in their independence training attitudes toward normal children than mothers with high achievement needs.

When the mothers are grouped into early and late independence training attitude groups on the basis of their attitudes toward deaf children, and these groups are equated for number of male and female children, it is found that the *n* Achievement scores of the mothers favoring late independence are significantly higher than those of mothers favoring early independence (Table 2, Line 2).

Table 2

Relationships Between Mothers' *n* Achievement and Independence Training Attitudes

Mothers Categorized on:	*n* Achievement Scores of Mothers Categorized as High and Low in Independence Training Attitudes			
	High	Low	*t* Ratio	*p*
1. Independence training attitudes toward normal child	*.22*[a]	*2.67*	*.87*	>.05
2. Independence training attitudes toward deaf child	*5.89*	*−1.22*	*2.89*	<.05
3. Discrepancies between attitudes toward deaf and normal children	*3.66*	.33	*1.10*	>.05

[a] Italicized figures are derived from analyses of sex-matched groups; nonitalicized figures are derived from analyses of groups not so matched.

The mothers were grouped according to the size of the discrepancies between their independence training attitudes toward deaf and toward normal children and the groups equated for number of male and female children in each. No significant differences in *n* Achievement scores were found between the mothers so grouped (Table 2, Line 3).

Discussion

Our results do not support the hypothesis that independence training attitudes toward deaf children are associated with the extent to which the social development of deaf children approaches the limits of their ability. However, independence training attitudes toward normal children were found to be associated with this variable. It may be that the measurement of attitudes under a "normal child" set provides a more valid estimate of independence training attitudes than does testing under a "deaf child" set, in that the responses of the mothers in the former situation are relatively uncontaminated by particular attitudes toward hearing handicaps. Our data suggest that in some indirect manner these attitudes toward normal children affect the behavior of the deaf children. It is to be noted that the direction of the relationship between children's behavior and the independence training attitudes is opposite to what was expected on the basis of Winterbottom's findings. A possible explanation for this may be that as a mother expects more and more from her normal hearing child, the deaf child stands in greater contrast to the normal sibling. Such contrast might operate on the deaf child by robbing him of the incentive to accomplish the developmental tasks, which compose the Vineland, through frustration and the inability to compete with the sibling. Thus discrepancies between the deaf child's performance and his ability develop.

An alternative explanation may be that the mother who encourages early independence in her hearing child is free to devote more of her interest and nurturance motives to the handicapped child. The increased interest may take the form of solicitousness, which would tend to reduce the opportunity and/or need for the deaf child to develop the kind of independence measured by the Vineland.

Rehabilitation workers sometimes take the view that an improvement in significant maternal attitudes toward the handicapped child, and thus ultimately improvement in the child's behavior, is contingent upon a reduction of emotionalized responses to the child and his handicap. Our data suggest, on the other hand, that the more important influences on the handicapped child are not the maternal attitudes toward the handicap but, rather, toward children in general.

The curvilinear relationship between achievement motivation in mothers and independence training attitudes toward normal children is an interesting one. The relationship indicates that highly motivated mothers expect independence either very early or very late. This may suggest that achievement motivation may interact with response dispositions as in Hullian drive theory to intensify response strength; thus, high drive mothers are found at the extremes of the distribution in independence training attitudes toward normal children.

However, this kind of relationship between motive and response is not found where the response is directed toward deaf children. Here achievement motivation is clearly associated with the failure to encourage early independence. It may be that the occurrence of a handicap in the family operates to focus maternal achievement needs, such that high *n* Achievement mothers perceive the handicap as a challenge. The overprotective response implied in the late independence training attitudes of high *n* Achievement mothers may be of such a nature that reduction of the drive to meet the challenge occurs, thus reinforcing protective attitudes. The increasing independence of a deaf child, on the other hand, may be perceived as a threat to the mother who views the handicap as a challenge which *she* must meet. Thus the encouragement of early independence in a deaf child would not be reinforced. Perhaps it is for reasons such as these that with reference to independence training attitudes toward a deaf child, high achievement motivation operates unidirectionally, rather than simply by intensifying early training or late training response dispositions, as was the case with attitudes toward normal children.

To summarize, the over-all picture suggests that high *n* Achievement mothers tend to overprotect their deaf child, or to wish to, but they may go to either extreme with their normal children. If they go to the extreme of favoring relatively early independence in normal children, the result may be a tendency for the deaf child to develop independence at a slower pace than his capacities would suggest.

Summary

This research was designed to test three hypotheses concerning the behavior of hearing handicapped children: (*a*) The extent to which a deaf child's social maturity level approaches the limits of his ability is related to the independence training attitudes of the child's mother; (*b*) The extent to which social maturity levels of deaf children approach their ability levels is related to the extent to which mothers' independence training attitudes toward deaf children differ from their attitudes toward normal children; and (*c*) Maternal independence training attitudes are associated with mothers' achievement motivation.

*S*s were 19 mothers and their deaf preschool children. Six TAT cards were administered to the mothers, following McClelland's standard instructions, and their stories scored for *n* Achievement according to McClelland's criteria. Mothers were also administered part of Winterbottom's Independence Training Questionnaire, once with the set to respond with reference to their deaf children, and once with a set to respond with reference to their normal hearing children.

Social maturity relative to intellectual ability was measured in the deaf children by the discrepancy between Merrill-Palmer and Vineland Social Maturity Scale scores.

The major results were as follows:

1. Attitudes favoring early independence in normal children were found in the mothers of deaf children who showed relatively large Merrill-Palmer discrepancies.

2. Independence training attitudes toward deaf children, and discrepancies between attitudes toward deaf and normal children were not associated with Merrill-Palmer—Vineland discrepancies.

3. Mothers with low achievement needs tended to be moderate in their independence training attitudes toward normal children, while high *n* Achievement mothers tended toward the extreme positions of favoring very early or very late independence.

4. High *n* Achievement mothers favored significantly later independence in deaf children than did low *n* Achievement mothers.

Received April 8, 1958.

REFERENCES

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953. P. 384.

Winterbottom, M. R. The relation of childhood training in independence to achievement motivation. *Dissertation Abstr.,* 1953, **13,** 440–441.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

POST-HOSPITAL ADJUSTMENT AS PREDICTED BY PSYCHIATRIC PATIENTS AND BY THEIR STAFF[1, 2]

JULIAN J. LASKY

VA Central NP Research Unit, Perry Point, Md.

GERALD L. HOVER, PHILIP A. SMITH, DAVID W. BOSTIAN

Veterans Administration Hospital, Ann Arbor, Michigan

STANLEY C. DUFFENDACK

Anglo-American Oil Company, Iran

AND CHARLES L. NORD

Nassau County Mental Hygiene Department, Mineola, New York

Psychologists have long been concerned with various aspects of the ability of people to judge others (Taft, 1955). Studies based on peer, co-worker, or buddy ratings, i.e., ratings of an individual by other members of his group, have been summarized, in part, by Hollander (1954a). The method rests on the assumption that peers, by virtue of their close association with one another, have opportunity to acquire a global fund of information about each other and that this information can be sampled by appropriate means.

Webb studied the reliability and validity (1955) and methodology (1956) of peer ratings. Wherry and Fryer (1949) found that buddy ratings were reliable and identified leadership potential among army officer candidates much earlier than did tactical officers and more accurately than did academic instructors.

Hollander (1956b) studied the length of time it took naval officer candidates to become well enough acquainted with each other before they could make valid inferences about future performance. Peer ratings made during orientation week predicted instructors' final (16th week) military aptitude grades. Willemin and Rosenberg (1957) observed that mental ability was not related to the ability to judge fellow soldiers among advanced infantry trainees.

In a study on morale, Webb and Hollander (1956) found peer ratings superior to both self-evaluations and to a survey technique. Willemin, Rosenberg, and White (1957) found that buddy ratings predicted the six areas (group loyalty, dependability, knowledge of the combat task, performance under stress, aggressiveness, and initiative) of the Combat Aptitude Scale better than did predictions based on personality, reasoning, perceptual speed, perception under distracting conditions, psychomotor, and Army Classification tests.

Thus, peer ratings have formed the basis for valid predictions regarding estimates of intelligence (Webb, 1955), morale (Webb et al., 1956), leadership (Wherry et al., 1949), success in aviation cadet training (Hollander: 1954a, 1954b, 1956a), performance in medical school (Kelly, 1957), and army combat effectiveness (Willemin, Birnbaum, & Rosenberg, 1957).

---

[1] We are deeply indebted to the patients, psychology trainees, and staff of the Ann Arbor VA Hospital Psychiatric Service who so generously gave their time and cooperation. This study was carried out between June, 1954 and August, 1956. E. Lowell Kelly of the University of Michigan was most helpful in reviewing the statistical analysis and reading the manuscript.

[2] A preliminary draft of this article was presented during the VA Day Program, American Psychological Association, Chicago, 1956.

Peer ratings have not been applied to the many problems of prediction and evaluation as related to mental patients. This study adapts the method of peer ratings to the study of prognosis in such patients. The purpose of this study was to investigate whether a group of open ward psychiatric patients from a VA general medical and surgical hospital could predict, by means of peer ratings, their own and each other's adjustment in four specific areas during the two-year period following their discharge from the hospital. The treatment staff also rated these patients by following the identical procedure which the patients used.

Method and Results

Predictions of the prospect of rehospitalization, and the post-hospital work, family, and health adjustment of all patients on the ward were made once each month for six months by every patient and staff member. Each *S* was individually interviewed and asked to check Yes, No, or Cannot Say after each patient's name for each of the following predictor questions. Each patient also rated himself.

1. Do you believe that this man will have to be hospitalized again for a neuropsychiatric condition within a period of two years following discharge?
2. Do you believe that this man will work steadily for the next two years after he leaves the hospital?
3. Do you believe that this man will have serious family trouble within the two-year period following discharge?
4. Do you believe that this man will enjoy good physical and mental health for the two years following discharge from the hospital?

The staff of 21 included: a staff psychiatrist; five psychiatric residents; three each of aides and nurses; two each of occupational therapists, recreation therapists, social workers, and psychologists; and a ward secretary.

Of the 36 patients, approximately 40% were neurotic, 20% schizophrenic, 25% had psychophysiologic reactions, and 15% were character disorders. Their median age was 31.5 years. Although patients were discharged from the hospital at varying intervals, each patient was included in three monthly surveys, on the average.

After two years, follow-up data were collected on all 36 patients. Of the 26 patients still residing in the state, 24 were studied at the hospital by a staff interviewer and a staff observer. One patient was seen by a prison psychiatrist and psychologist, who forwarded the necessary data, and the other was visited at his home by a Red Cross field worker. Complete data on the remaining 10 patients were obtained through the cooperation of the Red Cross, the VA, and the Army, and through correspondence with the patients or their relatives.

The post-hospital data were studied in terms of the following criteria. Each criterion was dichotomized, paralleling the form of the predictor questions.

Rehospitalization. Any inpatient hospitalization on any psychiatric ward during the two-year post-hospital period.

Work adjustment. A work index was devised to evaluate a patient's vocational adjustment:

$$\text{Work index} = \frac{\begin{array}{c}\text{time employed}\\ \text{since discharge}\end{array} + \begin{array}{c}\text{duration of longest job}\\ \text{held since discharge}\end{array}}{\text{total time since discharge}}$$

The work index value was doubled if the patient was employed and working at the time of the follow-up examination. The arbitrary weights of the variables comprising this work index yielded values which ranged from 0.0 to 4.0. A value of 3.0 or greater was set in advance as the criterion for satisfactory work adjustment. For example, a patient was judged to be making a satisfactory work adjustment if he was currently employed and had held one job at least 75% or more of the time since hospital discharge, or a patient who had held two jobs consecutively and had worked at them full time without any period of unemployment, etc. Two judges, using their own standards, dichotomized the post-hospital work history data independently and agreed with the categorization derived from the work index in 30 of 36 cases (83.3%). The work index was used as the sole criterion of work adjustment in computing correlations which appear later in this study.

Family adjustment. In order to meet the criterion of satisfactory family adjustment, the patient had to have a history free of the following since leaving the hospital: desertion, long separation, divorce, prolonged extramarital affairs, rejection of an acceptable family home, or acting out behavior (alcoholism, criminal or sexual assault) directly attributable to family conflict. Using this definition, two independent raters agreed on 30 of 36 cases (83.3%). The remaining differences of opinion were discussed and reconciled by the raters.

Health adjustment. A patient was defined as having good general health if any symptom(s) that he had manifested was not sufficiently disabling so as to seriously interfere with his over-all functioning as a citizen. Two independent judges agreed on 32 of 36 cases (88.9%). The four cases on which opinions differed were discussed and the differences were reconciled by the raters.

Two preliminary questions were asked of the data. What was the reliability of the ratings of the patient-peer group, and of the staff group? How closely did the patients agree with the staff in making their predictions? Each patient was given a Yes or No designation each month as long as he remained in the study on each of the four prediction questions. For example, a Yes for a patient during a particular month on the first question meant that the *majority* of the rating group (peers or staff) were of the opinion that month that this patient would be rehospitalized on a psychiatric ward within two years following his discharge from the hospital, etc. The average of all the biserial correlation coefficients between adjacent pairs of monthly group ratings over the six months' study period was .60 for the patient-peers and .77 for the staff. Both of these correlations were significant at the .01 level, and the difference between them was also significant at the .01 level. It is not too surprising that the staff ratings were more reliable than the patients' ratings in view of the high rate of turnover among the patients and the absence of personnel turnover among the staff. Biserial correlation coefficients computed each month for six months between the patients' distributions and the staff's distributions ranged between .57 to .82, all significant at the .01 level.

A tabulation of the criteria showed that about one third or 13 of the 36 patients had made an excellent adjustment in all four areas. About two thirds or 23 patients were making a satisfactory work adjustment, but two thirds had serious family trouble. One third or 12 patients had failed to make a satisfactory adjustment in all four areas: were rehospitalized, had not worked steadily, had had serious family trouble, and had not enjoyed good health.

After each patient was evaluated with respect to the criteria, the main task of comparing the predictors with the criteria followed, for each of the two rater groups and for each of the four areas studied. First, predictions based on each of six *monthly* surveys were compared, one at a time, with the criteria. Next, average ratings for each patient, based on three to six monthly surveys, depending on length of hospitalization, were computed, and these averaged ratings were compared with the criteria. These averaged ratings will be called *combined* ratings. Finally, the last ratings which a patient was given during the month before he was discharged from the hospital were compared with the criteria. These last ratings will be called *predischarge* ratings.

Chi square and Fisher's Exact Method were used to screen the predictive efficiency of the monthly, combined, and predischarge rating distributions. This procedure eliminated the individual monthly peer and staff ratings as predictors of post-hospital adjustment since none of the p values are significant. An inspection of chi squares for cells revealed the interesting tendency for patients to err on the side of being overly optimistic, while the staff erred by being too pessimistic regarding the anticipated adjustment of the patients. Computations based on the chi square "predischarge" cell values revealed that both patients and staff had an average accuracy of 70% in predicting adjustment in all four areas.

The main findings of this study are presented in Table 1, where it will be seen that 14 of the 16 correlation coefficients are significant at the .03 point. In nine instances, the correlations are significant beyond the .001 point. Staff raters were able to predict post-hospital adjustment at a level significantly greater than chance in all four areas. The patients predicted accurately for all ex-

Table 1

Relationships Between Majority Opinion Prediction Ratings and Corresponding Dichotomized Post-Hospital Adjustment Criteria for Patient Raters and for Staff Raters[a]

	Prediction Ratings			
	Combined		Predischarge	
Adjustment Criteria	Staff	Patients	Staff	Patients
I. Rehospitalization				
r_{bis}	.66	.58	.62	.49
σ_{est}	.14	.16	.15	.17
p	.001	.001	.001	.01
II. Work adjustment				
r_{bis}	.75	.36	.86	.39
σ_{est}	.12	.19	.09	.19
p	.001	.03	.001	.02
III. Family adjustment				
r_{bis}	.54	.78	.59	.84
σ_{est}	.16	.11	.16	.10
p	.001	.001	.001	.001
IV. Health adjustment				
r_{bis}	.41	.04	.43	.28
σ_{est}	.19	.22	.19	.21
p	.02	.42	.02	.09

[a] One-tailed test, $N = 36$.

cept the area of health adjustment. Among the three areas of adjustment that both groups could predict, the staff was significantly more accurate ($< .01$) in predicting work adjustment than the patients, but the patients were more accurate in predicting family adjustment ($< .05$).

Next, the predictive efficiency was evaluated for each patient regarding himself, each patient regarding other patients, each staff member as a rater, and each professional group represented on the staff.

Patient self-predictions did not reach the .05 point for any of the four areas of adjustment.

The predictions of only two to four patients on any one area of adjustment reached the .05 point. No patient was able to predict adjustment in all four areas consistently.

The most accurate predictors among individual staff raters included the ward secretary, an aide, a recreation worker, a nurse, a psychiatric resident, and the two psychologists. As far as individuals on the staff were concerned, ability to predict post-hospital adjustment of patients was not significantly related to professional background and training.

All professional groups on the staff were able to predict work adjustment, with the exception of the occupational therapists ($<$ .07). The two psychologists reached the .02 point of significance on all four areas. Social workers did less well than most other staff groups in predicting family adjustment, reaching a nonsignificant level of .14. Nurses, aides, resident psychiatrists, and recreation workers, as groups, predicted significantly at beyond the .03 point on rehospitalization, work, and family adjustment, but each of these staff groups missed on health adjustment.

Usually a statement regarding prognosis is made by the resident in charge of the case at the time of the patient's discharge. This condition was approximated statistically in this study by tabulating the predictions made by each resident for his own patients during the patients' last month of hospitalization. While none of the predictions made by the five residents, as individual raters, reached statistical significance, each resident's predictions were in the right direction. Since the number of patients per resident was small, these data were pooled. Pooled residents' ratings made during the patients' last month of hospitalization significantly predicted only one of the four areas, rehospitalization, at beyond the .01 point.

Other of the more usual prognostic variables were compared with the criteria. For this sample, diagnosis, length of treatment at this hospital, and age bore no significant relationship to post-hospital adjustment. Number of previous hospitalizations was a good predictor of rehospitalization ($<.01$) but was not a good predictor of work, family, or health adjustment. A rather unusual predictor was also evaluated. Biserial correlation coefficients were computed between weight of C-files[3] and criteria measures. The only criterion which the C-files did not predict at the .001 point was family adjustment, and this they predicted at the .05 point. The correlations ranged from .33 for family adjustment to .61 for rehospitalization, with the mean correlation being .52.

Discussion

Psychological tests, objective or projective, were not used in this study. Predictions were derived from human interactions. Judgments based on these interactions were obtained by asking for a Yes or No response from all of the people concerned.

During the first month or two of this study, some of the patients had difficulty accepting the fact that their opinions and predictions were regarded as being worthwhile. Although no study was made of this feature, the impression was that the attitudes the patients had toward the staff and toward themselves might have become more positive as a result of participating in this study.

The patient raters faced some sizeable obstacles. These included high rate of patient turnover and the inclusion of self-ratings, known in advance to possess little validity in this form, in the pool of patient ratings. Patients did not have access to the detailed information contained in previous and current hospital records, case histories, and test reports.

The staff raters also faced some obstacles. Apparently there were sources and kinds of information which were available to the patients but not to the staff. This is a problem area to be explored in its own right.

Methodologically speaking, neither rater group, on any one monthly rating period, could make valid predictions. However, the predictions attained significance if one distribution was constructed of ratings made during each patient's last month of hospitalization, or if all monthly surveys were combined into one distribution.

A methodological point which needs exploration is to investigate how much additional criterion information the raters could profitably use. The questions that were asked were simple and general. For example, "Will this man have serious family trouble?" The investigator could explain to one half of a future study population, "By serious family trouble we mean such things as this patient deserting his family, having long separations from his wife, obtaining a divorce, running around with other women or beating his wife." The other half of the group could be asked the simpler question.

It would be interesting to learn whether this general, rating by peers, approach could be applied in a psychiatric setting as it is done in a military setting; i.e., to collect data from a group of raters at one time, instead of individually. It is also interesting to speculate what might happen if the ward secretary or the patients, for example, through their Patient Council, were asked to collect data for this type of study.

As will be recalled, the patient's own self-predictions were unreliable. This was expected on the basis of other studies (Taft, 1955) involving self-ratings. Several unexpected events occurred, however. Various staff occupational groups were not able to predict the area(s) closest to their area of specialization, while their staff colleagues in other occupational groups were able to make these predictions. For example, social workers missed in predicting family adjustment. Occupational therapists missed in predicting job adjustment. Resident psychiatrists missed in predicting job, family, and health adjustment on their own assigned patients. Nurses, aides, and resident psychiatrists missed on the prediction of health adjustment. This raises a very interesting and broad question for further study. Do people lose some of their objectivity when they move close to themselves, or to their areas of professional competence?

---

[3] A C-file contains all the material on the veterans compensation claims and allowances, and VA military hospital discharge summaries.

Summary

Using a predischarge ratings by peers method, an open ward group of 36 psychiatric patients were able to predict individual patient's rehospitalization as well as their work, family, and health adjustment for a period of two years following hospitalization, with an average accuracy of 70%. The patients' staff of 21 members were able to predict future adjustment in the same four areas with the same average accuracy of 70%. The patients were more accurate ($< .05$) in predicting family adjustment, while the staff was more accurate ($< .01$) in predicting work adjustment. A biserial correlation coefficient of .84 was obtained between predischarge patient predictions of family adjustment and a two-year fellow-up criterion. A biserial r of .86 was obtained between predischarge staff predictions of work adjustment and a criterion.

As individuals, no patient was able to predict his own future adjustment nor was any patient able to successfully predict all four areas of adjustment for the other patients.

Ability to predict successfully among individual staff members or among the various professional subgroups was not related to professional specialization but, rather, seemed to be a function of the individual's ability to predict, per se. In the nature of a trend, it appeared that various professional groups were less able to predict the area of adjustment most closely related to their own area of specialization than were their staff colleagues.

Predictions based on single monthly ratings did not reach significance. Fourteen of 16 biserial correlation coefficients based on predischarge, or on combined (three to six monthly group ratings combined into one distribution) ratings were significant at beyond the .03 point.

Received April 10, 1958.

REFERENCES

Hollander, E. P. Buddy ratings: Military research and industrial applications. *Personn. Psychol.*, 1954, **7**, 385–393. (a)

Hollander, E. P. Peer nominations on leadership as a predictor of the pass-fail criterion in naval air training. *J. appl. Psychol.*, 1954, **38**, 150–153. (b)

Hollander, E. P. A better military rating system through peer ratings. *Res. Rev.*, 1956, July, 16–20. (a)

Hollander, E. P. Interpersonal exposure time as a determinant of the predictive utility of peer ratings. *Psychol. Rep.*, 1956, **2**, 445–448. (b)

Kelly, E. L. Multiple criteria of medical education and their implications for selection. In *Report of the fourth teaching institute*. Evanston, Ill.: Ass. of American Medical Colleges, 1957.

Taft, R. The ability to judge people. *Psychol. Bull.*, 1955, **52**, 1–23.

Webb, W. B. An analysis of the reliability and validity of "voting behavior" compared with individual judgments. *U. S. Naval Sch. Aviat. Med. Res. Rep.*, 1955, No. NM 001 109 101.05, 54-57.

Webb, W. B. A procedure for obtaining self-ratings and group ratings. *J. consult. Psychol.*, 1956, **20**, 233–236.

Webb, W. B., & Hollander, E. P. Comparison of three morale measures: A survey, pooled group judgments, and self-evaluations. *J. appl. Psychol.*, 1956, **40**, 17–20.

Wherry, R. H., & Fryer, D. H. Buddy ratings: Popularity contest or leadership criteria? *Personn. Psychol.*, 1949, **2**, 147–159.

Willemin, L. P., & Rosenberg, N. Evaluation of experimental reasoning, perceptual speed, and psychomotor tests for tryout as combat predictors. *Dept. Army, TAGO, Personn. Res. Bur. Tech. Res. Note 70,* 1957, PRB Project B-6-280-10.

Willemin, L. P., Birnbaum, A. H., & Rosenberg, N. Selection of experimental predictors for longitudinal validation in combat arms. *Dept. Army, TAGO, Personn. Res. Bur. Tech. Res. Note 72,* 1957, PRB Proj. B-6-280-10.

Willemin, L. P., Rosenberg, N., & White, R. K. Validation of potential combat predictors: Z I results for infantry. *Dept. Army, TAGO, Personn. Bur. Tech. Res. Note 76,* 1957, PRB Proj. B-6-280-11.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

CUES INFLUENCING JUDGMENT OF PERSONALITY CHARACTERISTICS[1]

ERNST G. BEIER AND JOHN STUMPF[2]
University of Utah

In judging another's personality characteristics one uses various cues, generally without ever recognizing the cues themselves. Different characteristics are likely to be judged from different cues, and some cues may afford more information than others. Here, accumulative information conveyed by successive presentation of four cues (voice, gestures, face, and social interaction) was evaluated with ratings on the following characteristics: intelligence, affective tone, sociability, productivity. Questions were asked with regard to relationship among ratings, the influence of additional cues on ratings, the significance of the introduction of previous ratings, and the comparison of "first impression" with "long acquaintance" ratings. Asch (1946) stressed the importance of the first judgment and its influence on succeeding judgments. Newcomb (1956) described the factors which enhanced interpersonal attraction as frequency of interactions, perception of reciprocated attraction, certain combinations of personality characteristics, and attitudinal agreement. Allport and Vernon (1933) studied perception of bodily expression and facial expression, giving attention to the emotional impact and accuracy of agreements in judgments. Wolff (1943) was concerned with agreement reached as to which expressions and behaviors of a given individual belong together. Again, the cues for each judgment were not of major concern. Secord, Dukes, and Bevan (1954), Asch (1946), and Newcomb (1956), in studies on interpersonal perception were more concerned with the specific cues comprising the formulation of judgment but investigated those cues within one event, for example, facial expression. Gage (1952) studied judgments of interests from expressive behavior through Kuder ratings given to the "judges" and "strangers" and discovered that the judges were fairly consistent in their ratings of five expressive tasks. While Gage's study appeared to be similar to the present study, it appeared that no study has used accumulative cues and the question of this consistency as a basis of research. In two investigations on the "Response to the human face as a standard stimulus" (Beier, Izard, Smock, & Tougas: 1953; 1957) "liking" and "disliking" ratings were used to investigate how judgments were formed in reaction to the human face. The present study was an expansion in two ways: instead of the face only, accumulative cues were introduced, and instead of a liking–disliking rating, a four-area rating was introduced.

Specifically, the present study investigated the following questions:

1. If four areas of personality characteristics (intelligence, affective tone, sociability, and productivity) were judged, would judgments remain constant as the judges (*J*) received cues cumulatively from observing the participant's (*O*) (*a*) voice, (*b*) gestures, (*c*) facial expression, and (*d*) interaction with the judges?

2. Would those judgments which did not remain stable become stable with some particular accumulation of cues, i.e., what was the minimum information the *J* needed in order to reach a stable decision?

3. Would the staggered introduction of ratings of the characteristics have any influence on the trend in which the characteristics were rated?

4. How did ratings of personality characteristics derived from first impressions differ from ratings made by individuals who knew *O*s for a long period of time?

---

[1] This research was supported in part by a U. S. Public Health Grant No. M 1926.

[2] The authors are indebted to Lyle E. Bourne for his assistance with the analysis of the data.

Table 1

Percentage of Student's Agreement on Selecting Words for a Given Characteristic

($N = 43$)

Characteristic	Positive		Negative	
Intelligence (I)	bright	100%	thoughtless	55%
	contemplative	100%	incapable	90%
	gifted	100%	stupid	100%
Affective Tone (A)	wholesome	100%	unhappy	100%
	content	100%	miserable	90%
	open	60%	hopeless	100%
Productivity (P)	industrious	100%	unreliable	60%
	practical	70%	distractible	55%
	organized	80%	slipshod	80%
Social (S)	cooperative	100%	revengeful	100%
	hospitable	100%	unfaithful	100%
	courteous	100%	standoffish	100%

Note.—These ratings came from an unpublished study done at Syracuse University. We are presently engaged in a new project where we have reason to believe that there are regional differences in such ratings.

Procedure

The task was (*a*) to develop an instrument to obtain ratings of the four personality characteristics, (*b*) to select *O*s as the subjects to be judged, (*c*) to devise a proper method of cue presentation, (*d*) to administer the instrument and cue presentation to the *J*s, and (*e*) to obtain ratings from a long acquaintance group.

1. Judgments were rated with words denoting four characteristics of personality: intelligence, affective tone, sociability, and productivity. The four characteristics themselves represented an arbitrary choice from among a large number of possible personality characteristics. The words had been gathered in the following manner: some 60 undergraduate students were given the task of writing down as many words as they could think of under each of the four characteristics. Each of the words was counted for the number of times it occurred under one heading. If it occurred at least 30 times (e.g., if at least 30 students wrote "bright" under "intelligence") it was used in a second alphabetical list of words. This list was presented to 43 undergraduate students who were given the task of assigning each of these words to one of the four characteristics. From the number of students who placed a given word into the selected characteristic, percentages of agreement were calculated for each word.

For the rating scales in the present study six words were selected for each of the four characteristics. There were three "positive" and three "negative" words for each characteristic. Each word was placed on a seven-point scale, the endpoints of which were called "least" and "most." The range of scores for the six words of one characteristic accordingly was 6–42. The group of words representing each of the four characteristics were ordered in such manner that four different booklets for administration were constructed. The first page would contain only one group of words, the second page two groups, the third page three groups, and the fourth page four groups. Each of the four booklets contained a different sequence of characteristics as can be seen in Table 2.

It was necessary to construct these different booklets to test whether the occurrence of a previous rating would influence the value of following ratings. Additionally, each of the four pages of all booklets contained a like–dislike scale, and space was left for comments and suggestions.

2. It was decided to use the junior janitorial help active at the University, some of whom planned to go to college and some did

not. Six junior janitors from the University of Utah were given the MMPI test. Three clinical psychologists using the method of paired comparison selected the four subjects most diverse in the characteristics to be rated. The *O*s who appeared most different from each other were selected for the study. They were male, between the ages of 16 to 19.

3. Each *O* was presented to a class of students. The various cues were introduced successively. For this purpose a stage was built with an attached curtain which could be lowered and raised. A screen shielded the *O* upon his entry from the view of the class. After the presentation of each cue *J*s rated *O* on the appropriate page of the booklet. The following cues were introduced: (*a*) *Voice*. Hidden by the curtain, *O* talked to the class for one minute. The talk was spontaneous since in the absence of reliable knowledge it was reasoned that a memorized talk would be too far removed from a life situation. The *O* was given the topic of marriage to discuss. Thus, this was not only a voice presentation but transmitted attitudes and opinions of the participant. (*b*) *Gestures*. The second cue introduced voice presentation with gestures. The curtain was raised so that arms, hands, and the upper torso between neck and belt was visible. Another topic, "father does not want to give car to son" was presented to the *O* with the instruction that he use gestures freely in this demonstration. Again the information contained more cues than could be obtained from gestures alone, and so the cue is somewhat misnamed. Perhaps some *O*s used this cue less spontaneously than others feeling that they had to "overact" to speak with their hands. This cue, on the surface, did not appear to have been a fortunate choice. (*c*) *Face*. Immediately following the ratings of gestures the curtain went up and *O*'s torso from the waist to the face was exposed to the *J*s. The *O* was told to look straight ahead and to refrain from any facial expression. After one minute exposure, the curtain was drawn and *J*s were instructed to make their third page ratings. (*d*) *Interaction*. The *O* was instructed to involve the class in a three-minute discussion of "cheating." The class was encouraged to respond. Most *O*s addressed the class for a minute and then asked questions. With this cue it is likely that attitudes and opinions were communicated in addition to the interaction. Generalization from this study will have to take this into consideration.

Table 2

Sequence of Characteristics in Each of Four Booklets

Booklet	Characteristics			
1	I	IA	IAS	IASP
2	A	AS	ASP	IASP
3	S	SP	ISP	IASP
4	P	IP	IAP	IASP

4. The 240 *J*s were male and female students from the University College and the School of Education of the University of Utah. The *J*s were given some over-all information on the project and asked to rate *O*s after the introduction of each cue. At the same time ratings on a like–dislike scale, and comments were requested.

5. The rating scale containing all the words was sent to four people whose names had been supplied by each *O*. It should be noted that this obviously would not represent a random group of long acquaintances but rather a highly selected group of people who would predictably react more positively toward *O*. The *E* told the long acquaintance *J*s that their ratings would be strictly confidential and would not influence the University's attitude toward *O*. All but one of the long acquaintances responded. These ratings were not thought of as *accuracy* ratings. The person who knew another person longer was not assumed to be more accurate in his observations though he did base his impression on a larger amount of information. The data was collected as a pilot study as the plan had been to develop eventually a design which would test accuracy data. Here one would learn about differences in ratings based on first impression made in a setting which could be expected to sponsor fairly objective attitudes (students participating in a scientific study) and ratings based on at least one year of acquaintance in a setting which would probably sponsor favorable attitudes. With the knowledge of these limitations in mind, the investigators examined the differences among the

Table 3

Means of the First Four Consecutive Ratings on Each of the Four Characteristics

	Sociability				Affective Tone				Intelligence				Productivity			
*O*s	Voice	Gestures	Face	Interaction	Voice	Gestures	Face	Interaction	Voice	Gestures	Face	Interaction	Voice	Gestures	Face	Interaction
1	31.2	31.0	32.6	33.3	35.3	32.5	35.2	36.0	29.1	29.6	33.4	36.7	32.7	29.7	33.5	35.7
2	31.0	30.8	31.46	31.7	26.73	30.0	31.3	32.7	26.9	26.2	27.4	27.1	25.2	27.8	27.0	26.7
3	30.5	30.2	30.8	32.6	26.8	29.7	26.93	29.73	24.7	25.5	27.7	28.3	25.1	26.5	26.6	26.4
4	31.73	31.0	31.8	31.7	29.6	29.8	29.0	31.0	28.0	29.7	30.8	32.7	27.2	26.7	27.9	28.9

Note.—*N* of 15 for each mean.

four characteristics as rated by the two groups.

Discussion of Results

Each of the four *O*s was judged by a group of 60 *J*s. Each of these groups was subdivided into four sections of 15. Each section was supplied with a different booklet (see Table 2). In this manner a total of 240 × 60 = 14,400 scores were obtained. This number was reduced to 2,400 by obtaining a sum of scores for each of the six words comprising a group. Three booklets had not been completed, and the scores for these booklets were averaged as all booklets were needed for the analysis of variance. Another deficiency in scale construction was discovered in that in 15 booklets one word was left out in one of the pages and, accordingly, only five words for the characteristic were averaged. A third error was introduced by the arrangement of the scoring sheet: a few semiattentive judges failed to maintain constancy and apparently exchanged the meaning of "most" and "least" occasionally.[3]

It should be remembered that the scores for each characteristic for each pamphlet were ordered in such a way that a group of *J*s rated this characteristic four times successively; another, three times successively; a third group, two times successively; and a fourth group, only once.

Three separate analyses of variance were performed on this data. The first is an analysis of the differences among four successive ratings on all characteristics for *O*s. The second is an analysis of the effects of previously available but unrated cues on *J*s first rating of *O*. The third measures the differences in final ratings after a varying number of previous ratings had been made.[4] In the following the basic description data will be presented.

The first analysis based on the means values of Table 3 was designed to answer the question of the change in the four ratings for all *O*s for the four succeeding cues. It can be concluded from the significant differences among the means of *O*s (F: 9.90, $p > .01$) that one of the assumptions of the test is met—that the *O*s are not alike in the critical characteristics. The finding that the interaction between *O*s and characteristics was not significant was interpreted to mean that while each *O* obtained different mean ratings, the characteristic profiles for each showed the same general trend on the consecutive ratings of each characteristic. *S*s/groups term (F: 12.36, $p > .01$) indicated that the *J*s showed individual differences in their ratings as it was hoped they would. The trials term (F: 31.24, $p > .01$) presents a rather interesting finding. As can be seen from the means there is not only a change across the ratings in the four cues, but the trend is constantly toward higher ratings. It appeared that either the accumulation of cues gave increasingly more positive information, or, otherwise the fact that the *J*s have *more* cues available—permitted the *J*s to express more favorable ratings on all characteristics. The $T \times O$ interaction (*F*: 13.35, $p > .01$) indicated that all ratings combined change for each *O* over the four trials. The .01 level of confidence was accepted for this study, so $T \times Ch$ (*F*: 2.05, $p > .05$) cannot be called significant. This would have to be interpreted that, while the ratings differ, the characteristics seem to have the same stimulus value on succeeding trials. By and large, there is little residual variance, indicating fairly high consistency among the ratings of *J*s.

---

[3] There was an attempt to correct these apparent slips, but only seven scores out of the total 14,400 were changed under rigorous criteria: (*a*) the score had to be extreme as, e.g., five "7" scores and one "1" score, (*b*) there was constancy on all other characteristics, (*c*) there was extreme constancy (7 or 1) on all following pages within this characteristic.

[4] The three tables of this analysis of variance have been filed with the American Documentation Institute. Order Document No. 5837, remitting $1.25 for microfilm or $1.25 for photocopies.

In these means (Table 4) it should be noted that the *J* has *seen* the previous cues in the sequence but had not had the opportunity to *rate* the same cue on the given characteristic, as ratings were introduced later on. The analysis deals principally with the hypothesis that previous available cues will have an effect on the *J*'s first rating of the *O*.

The *O* (*F*: 22.73, $p > .01$) and characteristic (*F*: 11.35, $p > .01$) sources of variance again indicate that *O*s differed significantly among themselves, supporting the selection process and that the characteristics differed from one another. The significant finding on cues (*F*: 11.40, $p > .01$) was interpreted to mean that the more cues the *J* has available to him before he attempts to rate, the higher will be the rating, a finding which is easily supported by an inspection of the means.

It becomes evident that the ratings did not become more favorable because of previous ratings, but primarily because of previous cues. The finding suggests that *J*s formed judgments of other people regardless of whether or not they were specifically asked to do so.

The significant $O \times Ch$ interaction (*F*: 3.61, $p > .01$) was interpreted as follows: while the *O*s have the same general profile on the consecutive ratings of each characteristic, the profiles differ when based on the staggered "first ratings" only. In other words, there is greater heterogeneity in the profiles of the *O*s based on data of first ratings from *J*s who had 0–3 previously unrated cues, and greater homogeneity when the cue was rated each time it was presented. Apparently, here, the very fact of rating had a constricting effect.

The $O \times Cu$ (*F*: 1.63, not significant) and $Ch \times Cu$ (*F*: 1.51, not significant) interactions support the finding of consistency of ratings.

The main question to be answered by the analysis of these means deals with the differences in final rating after various number of previous ratings have been made.

Support was found for the previous finding that *O*s were significantly different from each other (*F*: 34.02, $p > .01$) and that the characteristics differ (*F*: 10.42, $p > .01$). The finding of lack of significance between the means of previous judgments support the previously reported interpretation that the fact that a person rated a characteristic previously does not contribute to the differences

Table 4

Mean of Ratings for Each Characteristic on the Staggered Introduction of Cues

	Sociability				Affective Tone				Intelligence				Productivity			
Os	"First Ratings" on Voice	"First Ratings" on Gestures	"First Ratings" on Face	"First Ratings" on Interaction	"First Ratings" on Voice	"First Ratings" on Gestures	"First Ratings" on Face	"First Ratings" on Interaction	"First Ratings" on Voice	"First Ratings" on Gestures	"First Ratings" on Face	"First Ratings" on Interaction	"First Ratings" on Voice	"First Ratings" on Gestures	"First Ratings" on Face	"First Ratings" on Interaction
1	31.2	30.7	27.3	34.5	35.3	31.4	32.3	35.7	29.1	29.9	33.9	35.3	32.7	29.7	32.9	34.4
2	31.0	31.9	31.5	31.5	26.7	27.4	34.8	33.9	26.9	29.7	27.4	30.1	25.2	26.4	25.7	27.7
3	30.5	30.7	29.1	29.1	26.8	30.2	28.0	32.4	24.7	25.0	27.7	28.0	25.1	28.0	25.5	28.5
4	21.7	30.6	34.3	30.4	29.6	31.3	27.4	31.5	28.0	26.2	30.9	28.4	27.2	29.0	28.5	31.7

Note.—*N* of 15 for each mean.

Table 5

Means of Ratings of All Characteristics on the Final Cue: Interaction

	Sociability				Affective Tone				Intelligence				Productivity			
*O*s	With 3 Previous Ratings	With 2 Previous Ratings	With 1 Previous Rating	Without Previous Rating	With 3 Previous Ratings	With 2 Previous Ratings	With 1 Previous Rating	Without Previous Rating	With 3 Previous Ratings	With 2 Previous Ratings	With 1 Previous Rating	Without Previous Rating	With 3 Previous Ratings	With 2 Previous Ratings	With 1 Previous Rating	Without Previous Rating
1	33.3	34.1	32.7	34.5	36.0	34.5	33.4	35.7	36.7	34.6	35.5	35.3	35.7	34.1	35.0	34.4
2	31.7	34.8	33.7	31.5	32.7	32.4	32.3	33.9	27.1	26.4	27.5	30.1	26.7	26.7	28.8	27.7
3	32.6	29.8	30.4	29.1	29.7	31.5	29.8	32.4	28.3	27.5	29.2	28.0	26.4	29.9	28.0	28.5
4	31.7	32.0	35.7	30.4	31.0	34.5	31.5	31.5	32.7	29.5	30.5	28.4	28.9	29.6	29.0	31.7

Note.—*N* of 15 for each mean.

among ratings. Instead, it supports the view that the number of cues is the main contributing factor. The significant $O \times Ch$ interaction (*F:* 3.89, $p > .01$) indicated that the profiles of the four *O*s differ significantly in the final rating due to the varying number of previous ratings. In other words, while the *O*s have the same profiles on the consecutive ratings of each characteristic, the end result, i.e., the ratings on this last cue alone, do show difference as previously the staggered "first ratings" alone have shown differences. The *O*s appear similar in their general profile trend, but dissimilar in the values which make up their profiles, somewhat analogous to a paradigm of parallel wavy lines.

The lack of significance in $O \times R$ and $Ch \times R$ to all interactions with ratings gave support to the argument that previous cues, but not ratings, were the important determinator of the value of these ratings.

In Table 6 the trend generally indicated an increase in liking as more cues were presented, perhaps supporting the finding of increased favorable ratings. In two subjects (1 and 4) the means presented a clearly descending line, while in Subjects 1 and 3, only the final cue introduction shows an increase in "liking." The "gesture" cue seemed to contradict the trend most markedly, possibly because of its forced content, *O* was instructed to give emphasis to their talk by gestures. It might be desirable to select a more harmonious cut in the future as the forced behavior might bring intervening variables into the ratings. The comments made by the rating *J*s stimulated the investigators to think along the lines of a more "personalized" rating system in which the *J*s themselves would contribute their characteristics to be rated.

Finally, in a comparison of the ratings of the four *J*s of "long acquaintance" with the "first impression" *J*s, it appeared that three out of four subjects were underrated on all characteristics by the "first impression" *J*s. Only one subject was overrated by these *J*s, as compared to the "long acquaintance" *J*s. While the number of "long acquaintance" *J*s is too low to warrant an analysis of the data, it might be speculated that higher ratings of the "long acquaintance" *J*s may be at least partially due to the setting, where the 60 "first impression" *J*s had no stake in their judgment, while the "long acquaintance" *J*s were selected by the *O*s as people friendly to themselves and perhaps wanted to give as fa-

Table 6

Means of Ratings on Like-Dislike Scale by All *J*s on Four *O*s

($N = 60$)

Subjects	Voice	Gestures	Face	Interaction
1	3.12	3.57	3.03	2.43
2	3.66	3.37	2.92	2.75
3	3.63	3.73	3.78	2.23
4	3.50	3.35	3.05	2.85

Note.—Rating scale 1–7; most liked to least liked.

vorable ratings as possible. It would be interesting to discover to what extent the length of time of acquaintance really influenced the ratings and if the suggested tentative hypothesis, "Does familiarity breed appreciation?" has an empirical basis. Social class may very well be a factor in influencing judgments. Differences in social class between *J* and *O* may be a significant variable to be studied. The judges used here were above average intelligence. The intelligence of the *J*s as related to the intelligence of the *O*s should be another factor reflected in the ratings. Age is yet another factor in the present study; all *J*s were considerably older than the *O*s. These are questions which should be considered in future research.

Summary

In this study each of four *O*s were shown to groups of 60 *J*s for ratings of personality characteristics. The *O*s were presented under different cues. First the *J*s heard only *O*'s voice; second, *J*s also saw *O*'s gestures; third, *J*s saw *O*'s faces; and fourth, *J*s participated with *O* in a discussion. After each of the cue presentations, the *J* had to rate the *O* on one or more characteristics. The rating scales were composed of words connoting intelligence, affective tone, sociability, and productivity. They were devised in such a manner that each characteristic would occur once first, once second, once third, and once on the last cue only. In addition, ratings from four "long acquaintances" were obtained for each of the participants.

To the first question, which judgments remain constant under accumulation cue presentation, it was found that ratings for all cues and all participants generally showed an increase toward more favorable ratings.

The second question, whether or not the ratings level off with particular cut and particular characteristics had to be answered in the negative.

A third question, whether the staggered introduction of ratings on characteristics would influence the consequent ratings was answered in the negative. No significant influence was discovered.

The fourth question, concerned with difference between first impression ratings and long acquaintance ratings was answered by inspection only. It appeared that by and large the long acquaintance ratings were more favorable than first acquaintance ratings.

A few critical comments with regard to the study should be considered. One was the setting—the *J* worked in a group setting which may have minimized cue information gained on first impression. They also worked in a reasonably objective setting where they did not have to commit themselves, which is different from the "long acquaintance" who were likely favorably inclined toward the *O*s. The very conformity of the increase in favorability of the ratings for all cues and all *O*s was interpreted to mean that *J*s may have a common pattern to meet the stranger: "caution" and then "greater acceptance." It was thought likely that the group setting, the lack of involvement in the ratings, and the above average intelligence of the *J*s and the approximate homogeneity of their social class were more important determinants of the value of the ratings than individual difference of the *O*. Future research on these "setting cues" is recommended.[5]

Received April 11, 1958.

REFERENCES

Allport, G. W., & Vernon, P. E. *Studies in expressive movement*. New York: Macmillan, 1933.

Asch, S. E. Form impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.

Beier, E. G., Izard, C. E., Smock, D. C., & Tougas, R. R. Response to the human face as a standard stimulus. *J. consult. Psychol.*, 1953, **17**, 126–131.

Beier, E. G., Izard, C. E., Smock, D. C., & Tougas, R. R. Response to the human face as a standard stimulus: A re-examination. *J. consult. Psychol.*, 1957, **21**, 165–170.

Gage, N. L. Judging interests from expressive behavior. *Psychol. Monogr.*, 1952, **66**, No. 18 (Whole No. 350).

Newcomb, T. M. The prediction of interpersonal attraction. *Amer. Psychologist*, 1956, **11**, 575–582.

Secord, P. F., Dukes, W. F., & Bevan, W. Personalities in faces: I. An experiment in social perceiving. *Genet. Psychol. Monogr.*, 1954, **49**, 231–279.

University of California. Gough Adjective Checklist. Berkeley: Author, Inst. of Pers. Assessment and Res., 1955. (Mimeo.)

Wolff, W. *The expression of personality*. New York: Harper, 1943.

---

[5] At present, the data of a project by Beier and Korner, University of Utah, relating to influence of *J*'s intelligence on rating of *O*'s intelligence is being investigated.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

THE FACTORIAL DIMENSIONS OF THE MMPI[1]

GENE G. KASSEBAUM,[2] ARTHUR S. COUCH, AND PHILIP E. SLATER
Harvard University

The complex domain of personality has been subjected to numerous attempts at simplification through factor analytic techniques, but the results of these efforts have been somewhat disappointing to those who had hoped in this way to achieve major conceptual advances. Comparison of a sample of such studies often leaves the reader somewhat disheartened at the lack of continuity that he finds.

Careful reviews, such as that of Eysenck (1953), suggest, however, that the picture may be less black than it seems. Eysenck stresses the frequency with which the same two major personality factors have been isolated by investigators working with very different kinds of instruments and variables. He points out that similarities in results are often masked by variations in rotation positions, or by failure to carry out second-order analyses of oblique solutions. Even when solutions are similar, differences in the conceptualization of the factors imply a greater disparity between experiments than actually exists.

The first of the two recurrent factors observed by Eysenck suggests a dimension of ego-weakness vs. ego-strength, or, in Eysenck's terms, "general neuroticism" vs. "will." The second factor seems best defined by Jung's bipolar conception of introversion-extraversion (Jung, 1924; Eysenck, 1953, p. 318).

Factorial studies of the Minnesota Multiphasic Personality Inventory have until recently failed to approximate this general trend. A recent volume by Welsh and Dahlstrom (1956, p. 264) lists eight factorial studies of the MMPI, and Welsh comments:

> In almost all the studies there is essential agreement as to the loadings of two factors although the interpretations and the name assigned to these factors vary. One factor (usually the first) shows very high loadings on . . . (*Pt*) and . . . (*Sc*) with negative loading on scale *K* whenever it is used. A second factor shows fairly high loadings on the neurotic triad, especially . . . (*Hy*) and . . . (*D*), with negative loading frequently appearing on . . . (*Ma*).
>
> Unpublished studies have been carried out (by Welsh) on three different populations . . . all three studies confirm very clearly the factor loading patterns outlined above.

For the most part these studies were restricted in scope to the original clinical scales, and it was difficult with so few scales to define the factors clearly. More recent studies have profited from the development over the past decade of a large number of nonclinical scales. These provide wider coverage of the domain of personality than does the original battery, which was concerned primarily with the diagnosis of pathological phenomena. We will discuss these studies after first presenting our own research.

Procedure

This research grew out of a long-term laboratory study concerned with the prediction of interpersonal behavior in small groups. A preliminary goal of this research was to achieve a clarification of the dimensions of the MMPI in more general terms than the clinical emphasis of the test has heretofore allowed. For this purpose we included 19 nonclinical scales in our analysis in addition to the original 13 clinical and validity scales. The selection of these scales was based upon our interest in predicting certain aspects of interpersonal behavior, and this concern determined the domain of scales to be analyzed. Some scales have not been as fully validated as might be desired, while others, such as Gough's CPI scales (1957) are not fully represented in the

---

[1] This research was carried out under the direction of Robert F. Bales. The larger project of which the study is a part was facilitated by the Laboratory of Social Relations, Harvard. Funds were supplied by the Rockefeller Foundation.

[2] Now at Cornell University Medical College.

MMPI. Many of the scales have a number of items in common, and the intercorrelations among these scales are undoubtedly exaggerated thereby. Removing all overlapping items, however, almost obliterates some scales (Welsh et al., pp. 264–281), and the validity of these truncated scales is problematical. The authors felt that retention of overlapping

Table 1

Intercorrelations of 32 MMPI Scales

		L	*F*	*K*	*Hs*	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Pt*	*Sc*	*Ma*	*Si*	*Es*	*Ie*	*Lp*
1	*L*																
2	*F*	−05															
3	*K*	23	−42														
4	*Hs*	01	33	−33													
5	*D*	12	43	−32	52												
6	*Hy*	22	02	28	53	29											
7	*Pd*	−12	61	−37	41	45	21										
8	*Mf*	−11	31	−30	30	44	18	36									
9	*Pa*	04	26	−10	19	24	23	34	25								
10	*Pt*	−18	55	−65	47	64	04	63	52	43							
11	*Sc*	−23	68	−64	47	53	04	64	47	40	83						
12	*Ma*	−23	36	−37	32	04	07	51	20	00	39	46					
13	*Si*	00	49	−52	30	65	−17	42	41	21	63	59	−01				
14	*Es*	−08	−42	47	−50	−58	−15	−48	−41	−29	−66	−58	−19	−56			
15	*Ie*	−10	−48	47	−51	−54	−10	−49	−22	−17	−57	−55	−26	−50	55		
16	*Lp*	17	−53	65	−39	−62	12	−49	−48	−31	−79	−74	−21	−78	62	59	
17	*Ai*	04	−25	29	−16	−09	04	−14	15	−01	−22	−27	−30	−06	18	45	25
18	*Sy*	−04	−47	33	−20	−51	07	−32	−28	−29	−46	−47	09	−76	45	50	63
19	*Ac*	24	−46	55	−38	−30	−01	−52	−10	−17	−57	−66	−43	−35	31	43	51
20	*Re*	33	−35	36	−20	−07	09	−46	−13	−15	−40	−45	−52	−07	23	45	34
21	*Do*	−03	−33	24	−20	−32	06	−22	−18	−20	−46	−42	−12	−42	43	35	46
22	*Pr*	−21	59	−58	37	39	−08	58	27	07	66	67	58	45	−47	−66	−60
23	*St*	−10	−20	30	−19	−37	08	−24	−08	−12	−40	−33	04	−54	54	49	52
24	*Im*	−41	40	−52	36	17	03	61	26	18	62	67	64	15	−34	−41	−45
25	*Sp*	−01	−45	48	−27	−61	11	−39	−38	−32	−64	−58	−05	−80	61	57	71
26	*Fm*	14	16	−30	37	66	17	30	33	30	56	41	−02	53	−59	−43	−48
27	*Rp*	15	−56	64	−32	−50	16	−48	−38	−05	−60	−59	−28	−72	46	42	68
28	*R*	30	13	19	04	27	20	−09	04	06	−13	−13	−46	33	−04	−07	−03
29	*A*	−17	54	−70	38	54	−08	58	45	28	84	79	39	66	−67	−56	−79
30	*Dp*	−25	41	−67	43	58	−04	52	54	27	84	72	32	62	−65	−54	−77
31	*To*	24	−58	58	−39	−41	05	−58	−26	−08	−64	−67	−57	−46	47	64	60
32	*OI*	14	03	30	−12	−08	07	−14	08	−08	−22	−20	−25	−05	27	24	14

	Ai	*Sy*	*Ac*	*Re*	*Do*	*Pr*	*St*	*Im*	*Sp*	*Fm*	*Rp*	*R*	*A*	*Dp*	*To*	*OI*
Sy	09															
Ac	34	34														
Re	35	16	38													
Do	24	41	10	31												
Pr	−50	−36	−50	−52	−38											
St	18	52	12	12	42	−41										
Im	−28	−06	−60	−60	−14	59	−09									
Sp	15	71	29	17	50	−47	58	−20								
Fm	−02	−41	−18	−09	−42	27	−43	09	−58							
Rp	20	51	46	20	29	−57	32	−38	57	−31						
R	17	−37	13	40	03	−15	−14	−48	−28	09	−08					
A	−28	−45	−52	−41	−47	67	−47	55	−63	57	−60	−14				
Dp	−20	−41	−49	−40	−45	60	−41	53	−61	57	−56	−18	85			
To	46	34	47	51	37	−94	39	−59	48	−30	56	14	−66	−60		
OI	29	−02	14	34	22	−26	23	−30	17	−26	02	43	−29	−34	25	

items was the lesser of two evils, but in any case the close correspondence noted below between this study and a study using truncated or "prime" scales suggests that the practical significance of the problem posed by item overlap may have been exaggerated.

The 32 scales used are as follows:

L—Lie; *F*—Validity; *K*—Ego functioning; *Hs*—Hypochondriasis; *D*—Depression; *Hy*—Hysteria; *Pd*—Psychopathic Deviate; *Mf*—Masculine-feminine interest; *Pa*—Paranoia; *Pt*—Psychasthenia; *Sc*—Schizophrenia; *Ma*—Hypomania; *Si*—Social Introversion; *Es*—Barron's ego strength scale (1953)—predicts responses to psychotherapy; *Ie*—Gough's intellectual efficiency scale (1957)—a "nonintellectual intelligence test"; *Lp*—Oettel's leadership scale (1952); *Ai*—Gough's "achievement via independence" (1957)—predicts academic achievement in introductory psychology courses; *Sy*—Gough's sociability scale (1957); *Ac*—Gough's "achievement via conformance" (1957)—predicts academic achievement in high school; *Re*—Social responsibility scale (Gough, 1957); *Do*—Dominance scale (Gough, 1957); *Pr*—Gough's prejudice scale (1951); *St*—Gough's status scale (1948); *Im*—Impulsivity (Gough, 1957); *Sp*—Social presence (Gough, 1957); *Fm*—Hecht's feminine masochism scale (1950); *Rp*—McClelland's role-playing ability scale (1951); *R*—Welsh's factorial scale of "repression, denial" (Welsh & Dahlstrom, 1956, pp. 264–281); *A*—Welsh's factorial scale of "anxiety" (Welsh & Dahlstrom, 1956, pp. 264–281); *Dp*—Navran's dependency scale (1954); *To*—Tolerance scale (Gough, 1952); *O-I*—Originality potential (Gough, 1952).

The test was administered to 160 Harvard College freshman and scores obtained on these 32 scales. The mean scores for this sample approximate college norms for scales where such norms are available.

The 32 scales were intercorrelated and the matrix factor analyzed by means of the complete centroid method of Thurstone (1947). After the extraction of the third centroid, two criteria indicated the undesirability of further factoring: (*a*) the third factor accounted for only 5% of the total variance, and (*b*) the third factor residual matrix did not contain significant coefficients beyond the number expected by chance. The matrix was then rotated to simple orthogonal structure.

Results and Discussion

Table 1 shows the intercorrelation matrix, and reveals the generally high correlations between scales which other investigators have found.

Tables 2 and 3 reproduce the centroid and rotated loadings, respectively.

In Table 3 the main elements in the interpretation of Factor I are readily apparent. High saturations may be noted with regard to *Pt, Sc,* and Welsh's *A* (Anxiety), thus providing a close match with previous studies. High negative loadings are contributed by *Lp* (Leadership), *To* (Tolerance), *Es* (Ego-Strength), and *Ie* (Intellectual Efficiency). We have defined this factor as a dimension of *Ego Weakness vs. Ego Strength.* Factor I accounts for 39% of the total variance and 72% of the common factor variance.

Factor II exhibits high positive loadings on Welsh's *R* (Repression), *Re* (Social Responsibility) and *Si* (Social Introversion) and nega-

Table 2

Centroid Factor Loadings

Item No.		I	II	III	h^2
1	*L*	−.18	.41	−.06	.20
2	*F*	.65	−.07	−.20	.47
3	*K*	−.71	.17	.19	.57
4	*Hs*	.55	.13	.31	.42
5	*D*	.65	.48	.08	.66
6	*Hy*	.04	.21	.53	.33
7	*Pd*	.71	−.14	.17	.55
8	*Mf*	.48	.24	.27	.36
9	*Pa*	.34	.18	.25	.21
10	*Pt*	.91	.07	.14	.85
11	*Sc*	.89	−.10	.05	.80
12	*Ma*	.48	−.61	.18	.63
13	*Si*	.69	.46	−.38	.83
14	*Es*	−.73	−.30	−.09	63
15	*Ie*	−.72	−.05	.18	.55
16	*Lp*	−.85	−.20	.19	.80
17	*Ai*	−.36	.32	.28	.31
18	*Sy*	−.56	−.43	.38	.64
19	*Ac*	−.63	.29	−.03	.48
20	*Re*	−.53	.48	−.07	.52
21	*Do*	−.50	−.18	.09	.29
22	*Pr*	.71	−.28	−.26	.65
23	*St*	−.49	−.32	.30	.43
24	*Im*	.59	−.51	.23	.66
25	*Sp*	−.71	−.35	.28	.71
26	*Fm*	.55	.49	.17	.57
27	*Rp*	−.68	−.05	.24	.52
28	*R*	−.10	.60	−.34	.49
29	*A*	.88	.03	.02	.78
30	*Dp*	.85	.08	.12	.74
31	*To*	−.80	.32	.24	.80
32	*OI*	−.32	.24	−.14	.18
		$\Sigma a^2 = 12.62$	3.30	1.72	$\Sigma h^2 = 17.64$

Table 3

Rotated Factor Loadings

Item		I	II	III	h²
1	*L*	−.18	.40	.12	.21
2	*F*	.65	.02	−.21	.47
3	*K*	−.71	.07	.24	.57
4	*Hs*	.55	−.01	.34	.42
5	*D*	.65	.40	.28	.66
6	*Hy*	.04	−.03	.57	.33
7	*Pd*	.71	−.20	.09	.55
8	*Mf*	.48	.10	.35	.36
9	*Pa*	.34	.06	.30	.21
10	*Pt*	.91	.00	.16	.85
11	*Sc*	.89	−.11	.00	.80
12	*Ma*	.48	−.63	−.09	.64
13	*Si*	.69	.58	−.15	.84
14	*Es*	−.73	−.23	−.21	.63
15	*Ie*	−.72	−.12	.14	.55
16	*Lp*	−.85	−.26	.09	.80
17	*Ai*	−.36	.17	.39	.31
18	*Sy*	−.56	−.55	.16	.64
19	*Ac*	−.63	.28	.10	.49
20	*Re*	−.53	.46	.14	.51
21	*Do*	−.50	−.20	.01	.29
22	*Pr*	.71	−.14	−.35	.65
23	*St*	−.49	−.42	.14	.44
24	*Im*	.59	−.56	−.01	.66
25	*Sp*	−.71	−.44	.11	.71
26	*Fm*	.55	.37	.36	.57
27	*Rp*	−.68	−.15	.20	.52
28	*R*	−.10	.69	−.05	.49
29	*A*	.88	.02	.03	.78
30	*Dp*	.85	.02	.14	.74
31	*To*	−.80	.19	.35	.80
32	*OI*	−.32	.28	−.03	.18
		$\Sigma a^2 = 12.62$	3.38	1.66	$\Sigma h^2 = 17.66$

tive loadings on *Ma* and *Sy* (Sociability). The nature of this bipolarity suggested an interpretation of this factor in terms of *Introversion–Extraversion*. Factor II accounts for 10% of the total variance and 19% of the common factor variance.

The most important loading on Factor III is *Hy*, which has low saturations on I and II. None of the remaining loadings is of sufficient size or purity to be of substantial aid in the interpretation of the factor. It is worthy of note that the factor has loadings on all three members of the "neurotic triad," and is thus very similar to the so-called "neurotic factor" of some of the early studies. Other loadings, however, such as the positive loadings on *To* (Tolerance) and *Ai* (Achievement via independence) and the negative loading on *Pr* (Prejudice), suggest that the factor may be absorbing nonpathological elements from the clinical scales—an illustration of the hazards of factor interpretations based on clinical scales alone. We do not feel that a clear interpretation of this factor is possible at present, but have tentatively called it *Tender-Minded Sensitivity*. Because of its small size (it accounts for only 5% of the total variance and 9% of the common factor variance) and doubtful interpretation, Factor III was not included in the further analyses which are discussed below.

The Identification of Introversion

The first of these three factors, *Ego-Weakness vs. Ego-Strength*, has been isolated with such regularity in previous work as to be nonproblematic. Since Introversion–Extraversion has not heretofore been mentioned in connection with factorial interpretations of the MMPI, we shall discuss this identification in more detail.

The negative end of the factor, with high loadings on *Ma*, *Sy* (Sociability), *Im* (Impulsivity), *St* (Social Status) and *Sp* (Social Presence) may be readily identified with extraversion. The positive end, however, seems to contain a variety of scales whose common element is difficult to perceive. For the clustering of four of these positive loadings—*D*, *Si*, *Ac* (Achievement via conformance), and *O-I* (Originality potential)—we find considerable support from Eysenck, who quotes Guilford as saying: "It would seem that there is some basis (i.e., their high intercorrelation) for lumping together some characteristics bordering on seclusiveness with some implying a thinking person and still others that indicate depressed emotional tendencies and calling the resultant the introvert" (Eysenck, 1953, p. 104; Guilford & Guilford, 1939, p. 34).

The tendency of Factor II to split Ego-Weakness into "behavior disorders" (mania, impulsivity) and "neurotic disorders" (depression, feminine masochism) echoes studies cited by Eysenck (1953, pp. 60, 89–91, 125) as lending support to the psychoanalytic definition of introversion as superego predominance over id activities and extraversion as id predominance over superego activities.

Such a notion would help to explain the positive Factor II (Introversion) loading on *Re* (Social Responsibility).

No interpretation of Factor II will be acceptable, however, if it fails to take into account Welsh's *R* scale, which boasts the highest loading on Factor II and is virtually the only variable which is independent of Factor I. Inspection of the items in this scale suggests that Welsh's tentative label "repression, denial" is somewhat misleading, and that the terms "inhibition" or "constriction" might be more appropriate. The statements which are "denied" fall largely into two categories: (*a*) interest statements of a wide variety, such as "I like mechanics magazines," "I like dramatics," and "I think I would like the work of a building contractor"; and (*b*) expressions of hostile, gregarious, and excited feelings, such as, "At times I feel like smashing things," "At times I am full of energy," "I enjoy social gatherings just to be with people," and "I enjoy the excitement of a crowd" (Welsh & Dahlstrom, 1956, pp. 271–272). These statements would seem to appeal directly to the extravert, who is expressive and labile, and cathects new activities easily. The more cautious and inhibited introvert, who is slow to change, and seeks his satisfactions in his own inner life, would tend to reject statements of this kind (Eysenck, 1953, p. 21).

There is another characteristic of the *R* scale which is worthy of note, namely, that all of the items are counted when marked "false." This is equally true of the Lie scale, which also has a substantial loading on Factor II. Furthermore, when we examine the Mania and Depression scales (which have so often marked the two poles of this factor) we find that the Mania scale contains significantly more True items than the Depression Scale (chi square $= 17.64$, $p < .001$). This suggests the possibility that what we have labelled extraversion is associated with a general tendency to agree with any item whatever the content, while what we have called introversion involves a converse tendency to disagree or mark False.

This line of thought was prompted by an unpublished analysis by Couch and Keniston of a 360-item personality inventory, in which they found that the total number of positive (Agree) responses was a manifestation of a significant personality syndrome. This "tendency to agree" correlated positively with items that reflected a love of novelty and change, high stimulus reactivity, anal expressiveness, and a liking for enthusiastic but superficial involvement—all extraversive characteristics. Most prominent among negative correlates of "tendency to agree" was a scale developed by Couch and called "psychological inertia," which reflects a disposition to control impulsiveness (anal retentiveness), a slowness to change cathexes and a long secondary process in mental functioning—all introversive tendencies.[3]

This dimension seems highly related to our interpretation of Factor II as Introversion–Extraversion. A further support of this correspondence of dimensions is the correlation of $r = .65$, which Couch found between "tendency to agree" on the 360-item inventory and the tendency to mark True on the MMPI. This total true score on the MMPI correlates $r = -.75$ with Welsh's *R* scale, a finding which lends further credence to our generalization.

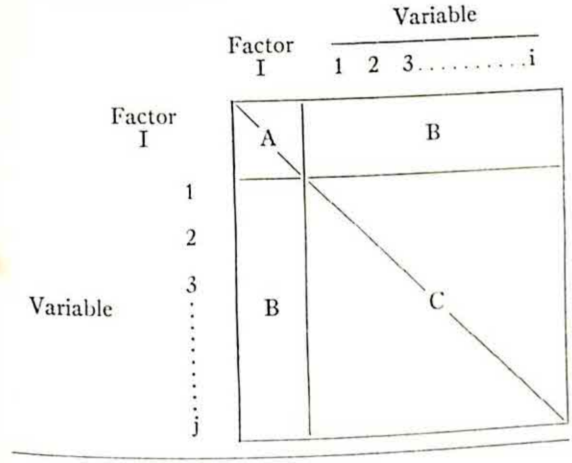
If this interpretation of Factor II is valid, it should be possible to give a new conceptualization to the vectors which lie midway between our orthogonal axes of Factors I and II. These four vectors should reflect Normal Introversion, Normal Extraversion, Disturbed

---

[3] It is rather interesting to find a detailed description of this aspect of introversion-extraversion in Freud, who, however, reserved the term "introversion" itself to a quite different characteristic which was associated with, rather than orthogonal to, neurosis. The Jungian conception of introversion is more closely approximated by what Freud called "adhesiveness of libido" (Freud, 1953, pp. 344–345). He refers to patients who "cannot make up their minds to detach libidinal cathexes from one object and displace them to another." He likens the analyst to a sculptor who with these cases works in hard stone, while with their antitheses molds soft clay or even, he adds despairingly, writes on water, due to the ease with which new cathexes are abandoned. In the adhesive patients he sees an "exhaustion of the capacity for change and development" and a "psychical inertia" quite independent of normal resistance (a distinction relevant to our discussion of the repression scale). Freud does not attempt to account for these types in psychoanalytic terms but attributes them to fundamental differences in the rhythm of psychical development.

Table 4

Theoretical Matrix for Pooling Square Technique



Introversion, and Disturbed Extraversion. But before examining the factor space in detail, it is necessary to discuss briefly the techniques we used in generalizing our analysis.

The Construction of Factorial Tests

We first wished to obtain reliable measures of the MMPI factorial dimensions, using the knowledge of the factor space gained from the factor analysis. For this purpose the pooling square technique (Thompson, 1949) is perhaps the simplest and most useful. It allows one to determine the hypothetical loadings, on each factor, of the sum of any group of variables included in the proposed measure. If a matrix is constructed as shown in Table 4, with A = the theoretical communality of the factor in question (1.00), B = the factor loadings of the several variables which are candidates for inclusion as measures of the factor, and C = the intercorrelations of these variables (obtained from the original correlation matrix), then the correlation of this set of variables (equivalent to an average of standard scores)[4] with the theoretically defined factor (denoted $A_{\hat{I}I}$) is

$$A_{\hat{I}I} = \frac{B}{\sqrt{AC}}$$

but in this case A = 1.0, thus

$$A_{\hat{I}I} = \frac{\sum a_i}{\sqrt{\sum r_{ij}}}$$

In this way it is possible to assess the loading of the entire scale on each of the dimensions of the factor matrix. The selection of variables to be included in the scale attempts to maximize the communality of the new test, and minimize the saturation on other factors by equalizing their positive and negative loadings.

The Factor Fusion Approach

We have long been of the opinion that if factor analysis is to make significant contributions to personality theory it is imperative to attempt to conceptualize the entire variable domain. For this reason, it is our practice to define and measure those vectors lying midway between the primary reference axes. In other words, once having established the primary reference axes by the usual simple structure criterion, we rotate the axes 45 degrees and repeat the procedure of defining the factors and constructing factor tests by the pooling square technique.

We have adopted the practice of calling these new factors "factor fusions," inasmuch as they may be conceived as combinations of the primary reference axes. To do so is not to overlook the fact that the decision as to which factors are to be considered the primary reference axes is a completely arbitrary one. It is our view that ideally nothing other than theoretical relevance should fix the primary axes in any particular position.

Although this approach is something of a departure from usual factor analytic procedure, the view that one should be able to measure any vector within the factor space was recently stated by Kuder (1954) and reflects a viewpoint that is becoming more widespread. The work of Leary (1957) and his associates, for example, is based upon a two-dimensional system of interpersonal behavior in which 16 equidistant vectors are measured independently. Our practice of defining and measuring those vectors lying midway between the primary axes is a first step in this direction. We feel that it has two advantages not realized in either standard orthogonal or oblique solutions.

---

[4] When the variable domain involves heterogeneous variances, as in the present study, it is necessary to convert all component scale scores to standard scores prior to computation of the new factor test score.

The first advantage is conceptual clarity. A better understanding of the nature of the fixed position factors is produced when they are examined in relation to other dimensions in the factor space. It is often particularly profitable to test the validity of the conceptualization of the original factors by obtaining predictions (made on the basis of these conceptualizations alone) of the nature of the factor fusions before the rotation is made.

The second advantage is the increased probability of a continuity of studies. The arbitrary placing of axes according to the simple structure criterion maximizes apparent diversity of findings if fusion positions are not also interpreted. This diversity is often due to such accidents as the selection of variables to be included in the analysis. For example, if Investigator A includes more tests which lie on Axes I and II, while Investigator B includes more tests which lie on Axes I′ and II′, their simple structure solutions will lie 45° from one another. An impression of great disparity will be created when in reality an almost perfect replication has been achieved.

The Fusion Factors

The fusion factor whose positive end lies between the positive ends of Factors I and II we will call Fusion Factor A. As can be inferred from Table 3, its highest and purest loadings are on *Si, D,* and *Fm* (Feminine Masochism) on the positive end, and *Sp* (Social Presence), *Sy* (Sociability), and *St* (Social Status) on the negative end. We have tentatively labelled this factor *Social Withdrawal vs. Social Participation.*

The fusion factor whose positive end lies between the positive end of Factor I and the negative end of Factor II we will call Fusion Factor B. Its highest and purest loadings are on *Im* (Impulsivity) and *Ma* on the positive end, and *Re* (Social Responsibility) and *Ac* (Achievement via conformance) on the negative. Originality (*O-I*) and *Ai* (Achievement via independence) also lie along the negative end of this axis but have more modest loadings. We have tentatively labelled this fusion factor *Impulsivity vs. Intellectual Control.*

Table 5 presents the pooling square estimated loadings of the constructed tests for Factors I and II and Fusion Factors A and

Table 5

Estimated Loadings on MMPI Factor Tests and Fusion Factor Tests

		I	II	III	h^2
Factor I	Ego-Weakness vs. Ego-Strength Psychasthenic, Schizophrenic, Anxiety, Dependency and Psychopathic Scales, (−) Leadership, (−) Tolerance, (−) Ego-Strength, (−) K-Suppressor, and (−) Intellectual Efficiency Scales	.99	.08	−.02	.98
Factor II	Introversion vs. Extraversion Repression-Denial, Social Responsibility, Social Introversion Scales, (−) Hypomania, (−) Sociability Scales	.04	.90	−.04	.81
		A	B	C	h^2
Fusion Factor A (I+II+)	Social Withdrawal vs. Social Participation Social Introversion, Depression, Feminine Masochism Scales, (−) Sociability, (−) Social Presence, (−) Social Status Scales	.94	.13	.02	.91
Fusion Factor B (I+II−)	Impulsivity vs. Intellectual Control Impulsivity, Hypomania Scales, (−) Social Responsibility, (−) Achievement via Conformance Scales	.09	.91	−.10	.85

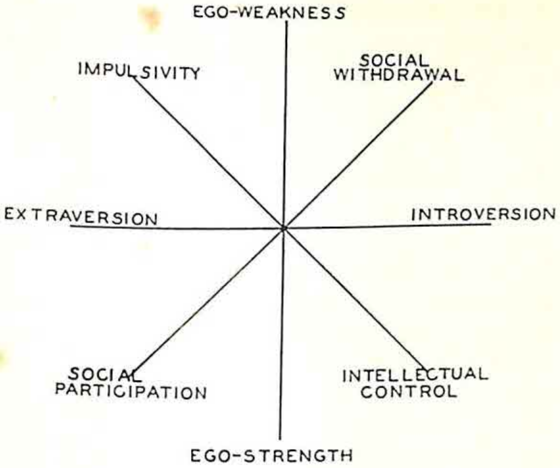


Fig. 1. Factors and factor fusions.

B, along with the scales of which the tests are composed.

Summary of Factor Interpretations

In order that the factors and factor fusions may be visualized in relation to one another, a graphical presentation of the four axes is shown in Fig. 1, with the labels we have chosen. This should help to clarify the summary description of the four dimensions which follow.

Factor I: A dimension of ego-weakness vs. ego-strength. Maladjustment, "general neuroticism," anxiety, dependency, and a tendency toward personality and behavior disorders, as opposed to healthy nonanxious personality, absence of disorders, and a tendency toward leadership and intellectual ability.

Factor II: A dimension of introversion vs. extraversion. High "psychological inertia" or "adhesiveness of libido" as opposed to ease and lability of object cathexis. Cautious, controlled, inhibited, intellectual, skeptical, intrapunitive personality vs. adventurous, impulsive, expressive, gregarious, naive, extrapunitive personality. In psychoanalytic terms this dimension expresses the superego-id balance within the personality, while Factor I expresses the strength of the ego.

Fusion Factor A: A dimension of social withdrawal vs. social participation. Shy, seclusive, submissive, guilty, depressed, masochistic personality, as opposed to poised, sociable, dominant, confident, and spontaneous personality.

Fusion Factor B: A dimension of impulsivity vs. intellectual control. Impulsive, manic, excitable, erratic, uncontrolled, delinquent personality as opposed to responsible, controlled, intellectual, stable, original, intraceptive, insightful personality.

It might be added that in psychoanalytic terms, Factor A seems to suggest the ego overwhelmed by a harsh superego (Depression), while Fusion Factor B suggests the ego overwhelmed by impulses from the id (Mania). Note, too, that Social Withdrawal and Intellectual Control, the two vectors adjacent to Introversion, provide a representation of the distinction often drawn between "social introversion" and "thinking introversion."

It should be noted at this point that these interpretations of the factors are based on a logical-inductive approach, which assumes the empirical validity of the MMPI scales. Our conceptions of the major dimensions are at present theoretical hypotheses, based on the inductive implications of the factor pattern. It would require further research involving psychological studies of subjects high, middle, and low on the factor scales to give more direct empirical support to these conclusions. Furthermore, the use of these major dimensional scales of the MMPI in clinical practice would add considerably to a clarification of the underlying dispositions that are being measured.

Comparisons with Other Studies

As suggested above, the strongest support for our factor interpretations comes from Eysenck (1953), who found repeatedly in his review of factorial studies of personality that the bulk of the common factor variance in these studies was accounted for by a general factor of neuroticism or ego-weakness, and a bipolar factor of introversion–extraversion. That this same structure should also emerge from a study of the MMPI is encouraging to those hopeful of isolating stable personality dimensions of widespread applicability. Studies of the Guilford inventory by Lovell (1945) and North (1949) seem particularly similar to ours.

We feel, however, that the isolation of our fusion dimensions adds most significantly to the struggle for factorial continuity. Some of the studies cited by Eysenck, particularly those of Goldman-Eisler (1951; Eysenck, 1953, pp. 123–126), seem to fit poorly into his two-dimensional framework. Although Eysenck sees a resemblance between Goldman's second factor of oral optimism–oral pessimism and extraversion-introversion, Goldman's first factor, defined by impulsion vs. deliberation, is hardly one of neuroticism or ego-weakness. It now becomes clear, however, that Goldman's factors very closely approximate our fusion factors, while her 45° vectors, though quite poorly defined, seem to approach our Factors I and II. Impulsion vs. deliberation corresponds to our *Impulsivity vs. Intellectual Control,* while oral pessimism (endocathexis, aloofness) vs. oral optimism (exocathexis, nurturance) corresponds to our *Social Withdrawal vs. Social Participation.* Eysenck (1953), for example, defines oral optimism as "distinguished by . . . generosity, bright and sociable social conduct, accessibility to new ideas and ambition accompanied by sanguine expectation," and oral pessimism in terms of "moods of depression and attitudes of withdrawal" (p. 124).

We should like to suggest the hypothesis that an orthogonal solution of a factor analysis of any personality inventory or battery of trait ratings of sufficient size and complexity, and drawn from a normal population, will produce a structure in which all eight 45° vectors of the two largest factors will be essentially similar to those we have identified. We feel that deviations from this structure are due only to peculiarities of samples, insufficiently "covered" domains, and the use of oblique solutions.

This hypothesis derives considerable support from an examination of the factorial studies of the MMPI. While a full review of this literature is not appropriate here, a brief survey should serve to illustrate the point that the more MMPI scales included in a factorial study of that instrument, the more closely the results approximate ours.

All of these studies have found a first factor similar to ours, but in the early studies, in which only the clinical scales were included, no factor emerged which could conceivably be conceptualized as introversion–extraversion. Hints were present only in the form of a small factor loaded on *Ma, D,* or both. The studies of Cook and Wherry (1950), and of Wheeler, Little, and Lehner (1951) fall into this category.

Tyler's study (1951), which included *Pr, Re, Do, St,* and *Ac* in addition to the clinical scales, falls into an intermediate category. His Factor III (Social Aggression and Friction), with positive loadings on *Ma, Pd,* and *Pa,* and negative loadings on *Si* and *Re,* resembles a mirror image of our Factor II.

Studies by Welsh and his associates (1956, pp. 264–281) also fall into this category. After a careful examination of previous factorial studies, Welsh constructed a number of special scales, including a set of revised clinical scales from which all item overlap had been eliminated. In addition, the *A* scale was constructed from items which differentiated extreme scorers on a scale of general maladjustment (*Gm*). The *R* scale was developed in a similar manner from a special "prime" (no item overlap) *D* scale. Welsh then factor analyzed this group of special scales and extracted three factors. No rotations were carried out. The first factor was loaded .99 on *A* and the second .88 on *R*.

Our analysis essentially replicates Welsh's findings. *A* is loaded .88 on Factor I, while *R* is loaded .69 on Factor II. Both the high loadings and the orthogonality of his scales are reflected in our analysis.[5]

The analysis which most clearly supports our findings is an unpublished study by Jerome Fisher (1957).[6] Fisher's variables include 23 MMPI scales and four additional measures. Of the five factors extracted, the second and fourth are heavily influenced by the non-MMPI variables. The first factor seems to reflect the familiar dimension of general ego-weakness. The third factor is the most interesting from our viewpoint, with high positive loadings on *Ma*, and high negative loadings on *R*, *Si*, and *D*. Other variables which helped to define our Factor II, such as *Re*, *Sy*, *Sp*, and *Im*, were not included in his analysis, but the correspondence is nonetheless striking.

---

[5] It might also be noted that if Welsh's second and third factors are rotated, his third factor also bears a remarkable resemblance to ours. The few differences existing between these two analyses are undoubtedly due in large part to his use of "prime" scales.

[6] We are grateful to Jerome Fisher for the opportunity to inspect his analysis in a dittoed progress report entitled, "An empirical study of the relation of physical disease to body-object cathexis."

That previous investigators have never conceptualized MMPI factors in terms of introversion–extraversion seems to be due in large part to their heavy dependence upon the original 13 scales. Another important consideration, however, is the fact that MMPI researchers have had little success in constructing scales unsaturated with Factor I, due presumably to the strong pathological emphasis of the instrument. Many researchers, moreover, have concentrated on the seclusive more than the intellectual and reflective aspects of introversion, and least of all on the central tendency of this complex trait.

It should be noted that none of the theories regarding typical pathological expressions of introversion and extraversion are borne out in this analysis. We do not find Factor II separating the cycloid from the schizoid, nor the hysteric from the psychasthenic. If we combine *Ma* and *D* by the pooling square method we find all four relevant measures ($Ma + D$, Sc, Pt, Hy) cluster obstinately around the zero point of this dimension. In the absence, however, of either theoretical agreement or experimental confirmation in the past regarding these relationships (Eysenck, 1953, p. 160), we feel justified in leaving the question open for the moment.

Summary

A factor analysis of 32 scales from the MMPI produced three factors, the first two of which were called Ego-Weakness vs. Ego-Strength, and Introversion–Extraversion, respectively. In discussing conceptualization of the second factor it was suggested that the trait extraversion may be expressed in part by a general tendency to agree with all items regardless of content.

We have argued strongly that continuity in factor analytic research is completely dependent upon interpretation and measurement of the entire factor space, and a simple procedure for achieving this goal was outlined and applied to this study. A method of constructing reliable and factorially "pure" factor tests was also described.

A brief review of other factorial studies of personality revealed greater continuity than is usually apparent when interpretation of vectors other than the primary reference axes is neglected, and adherence to an arbitrary simple structure criterion is maintained. It was hypothesized that the two-dimensional structure which emerged in our analysis will be found in any large-scale factorial study of the personality domain, provided the study is of sufficient scope and the entire factor space interpreted.

Received April 29, 1958.

REFERENCES

Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.

Cook, E. B., & Wherry, R. J. A factor analysis of MMPI and aptitude test data. *J. appl. Psychol.*, 1950, **34**, 260–265.

Eysenck, H. J. *The structure of human personality.* New York: Wiley, 1953.

Fisher, J. An empirical study of the relation of physical disease to body-object cathexis. Unpublished manuscript, VA Hospital, 42nd & Clement, San Francisco, Calif., 1957.

Freud, S. *Collected papers.* Vol. V. London: Hogarth, 1953.

Goldman-Eisler, F. The problem of orality and of its origin in early childhood. *J. ment. Sci.*, 1951, **97**, 765–781.

Gough, H. G. A new dimension of status: I. Development of a personality scale. *Amer. sociol. Rev.*, 1948, **13**, 401–409.

Gough, H. G. Studies of social intolerance: II. A personality scale for anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 247–255.

Gough, H. G. *Manual for the California Psychological Inventory.* Palo Alto, Calif.: Consulting Psych. Press, 1957.

Gough, H. G. Special keys developed for the MMPI. Unpublished manuscript, Univer. of California, Berkeley, 1952.

Guilford, J. P., & Guilford, Ruth B. Personality factors D, R, T, and A. *J. abnorm. soc. Psychol.*, 1939, **34**, 21–36.

Hecht, Shirly. An investigation into the psychology of masochism. Unpublished doctoral dissertation, Univer. of California, Berkeley, 1950.

Jung, C. G. *Psychological types.* New York: Harcourt, Brace, 1924.

KUDER, F. G. Expected developments in interest and personality inventories. *Educ. psychol. Measmt,* 1954, **14**, 265–271.

LEARY, T. *Interpersonal diagnosis of personality.* New York: Ronald, 1957.

LOVELL, C. A study of the factor structure of thirteen personality variables. *Educ. psychol. Measmt,* 1945, **5**, 335–350.

MCCLELLAND, W. A. A preliminary test of role-playing ability. *J. consult. Psychol.,* 1951, **15**, 102–108.

NAVRAN, L. A rationally derived MMPI scale to measure dependence. *J. consult. Psychol.,* 1954, **18**, 192.

NORTH, R. D. An analysis of the personality dimensions of introversion-extraversion. *J. Pers.,* 1949, **17**, 352–367.

OETTEL, A. Leadership: A psychological analysis. Unpublished doctoral dissertation, Univer. of California, 1952.

THOMPSON, G. *The factorial analysis of human ability.* London: Houghton Mifflin, 1949.

THURSTONE, L. L. *Multiple factor analysis.* Chicago: Univer. Chicago Press, 1947.

TYLER, F. T. A factorial analysis of fifteen MMPI scales. *J. consult. Psychol.,* 1951, **15**, 451–456.

WELSH, G. S., & DAHLSTROM, W. G. *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956.

WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.,* 1951, **15**, 134–141.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

THE PROBLEM OF INTERMINABILITY IN OUTPATIENT PSYCHOTHERAPY

DONALD R. STIEPER AND DANIEL N. WIENER
Veterans Administration Mental Hygienic Clinic, St. Paul, Minnesota

Termination has rarely been studied as a dynamically important and necessary event in psychotherapy or counseling. The vagueness with which most texts on psychotherapeutic techniques deal with it attests to the unstable nature of the conceptualization behind acts of therapeutic termination. The problem seems to center around appropriate criteria for terminating clients.

The literature on the subject of termination splits itself into three major classes: the first tends to deal in generalities with the necessary conditions for successful termination. For example: Weigert (1952) states that "the resolution of the countertransference is presented as a means of determining the approaching termination of psychoanalysis. . . ." Masserman (1955) concludes that "the completion of psychotherapy . . . is indicated not by what the patient says but by (*a*) how he feels, (*b*) how well he functions, and (*c*) what his prospects for continued improvement without further therapy are reasonably judged to be."

A second group of papers on termination deals largely with empirical data on the relationships between length of treatment and improvement. Notable among these are the studies reported by Rogers and Dymond (1954) in which Seeman (Chap. VII) reports that "there is a trend in favor of higher success ratings for longer cases" (at least twenty interviews). The major criterion for termination in the counseling program at that time was the client's decision to drop out of therapy. Myers and Auld (1955) corroborate this finding, reporting that there is a greater frequency of success ratings by therapists for patients in treatment for "twenty interviews or more . . . ," than among short-term patients. As in many studies of this type, the category "twenty or more" leaves the upper limit much more poorly defined than the lower one.

A third group of studies attempts descriptively to relate other variables to length of time in therapy. Middle-class patients tend to stay in therapy significantly longer than the lower-class patients (Imber, Nash, & Stone, 1955; Winder & Hersko, 1955). Therapists tend to keep patients in treatment longer when they initially perceive them to be healthier (Conrad, 1954). Inexperienced therapists tend to have greater difficulties terminating patients than experienced therapists (Auld & Myers, 1954; Myers & Auld, 1955; Myers & Schaffer, 1954; Schaffer & Myers, 1954).

In several studies (Kotkov & Meadow, 1953; Rogers, Knauss, & Hammond, 1951; Schaffer et al., 1954), duration of therapy is considered the criterion for success, on the assumption that successful therapy requires long periods of intensive work. Again, the upper limits of this relationship remain unconsidered, so that the patient in interminable treatment is left in limbo.

The most notable early attempt to deal with termination in a systematic, definitive fashion was made by Otto Rank (1929), who postulated the entire therapeutic process to be a "preparation for separation," a synthetic re-enactment of the original experience of birth trauma, or physical separation from the mother. Characteristically, the point of termination was set early in treatment, and the patient gradually worked toward the experience of successful termination.

Freud (1953) himself experimented with this approach in the attempt to guard against what he described as "interminable analyses." He concluded that "the measure is effective, provided that one hits the right time at which to employ it." He added, however, that ". . . while the force of the threat will have the effect of bringing part of the material to light, another part will be held back and become buried, as it were, and will be lost to our therapeutic efforts."

The language of therapy today has provided a convenient rationale for seemingly interminable therapy cases. The term "maintenance therapy" implies that, for selected patients, the ability to function without the continued support of their therapists is constricted. This concept stems, in part, from Federn (1953) and other "ego-psychologists" who hypothesized that in seriously ill (schizophrenic) patients, there exists a genetically weak ego with highly vulnerable ego-boundaries, who cannot function without the aid of another supportive ego—such as that of relative, therapist, or nurse.

The questions remain: Is indefinitely protracted (interminable) therapy the treatment of choice in any cases? Is it purposive, efficient, therapeutic? Or is it the result of poorly conceived therapeutic goals based, in part, on unsubstantiated theory?

This paper presents the results of a small investigation into the circumstances surrounding protracted (and seemingly interminable) psychotherapy with four groups of patients in one outpatient setting.

Procedure and Results

This study was designed to assess the results of brief, and prolonged, periods of psychotherapy in an outpatient setting. The total sample consisted of 66 patients, subdivided into the following four groups:

Group A: 15 white, male veterans currently being seen monthly by practitioners outside the Veterans Administration on fee basis. Mean length of time in therapy is 7.9 years, ranging from 3 years 6 months to 10 years.

Group B: 18 white, male veterans currently being seen by therapists on the staff of the VA Mental Hygiene Clinic. They also are being seen on an average of once monthly. Mean length of time in therapy is 5.3 years, ranging from 2 years 8 months, to 7 years, 11 months.

Group C: 15 white, male veterans currently being seen by therapists on the staff of the VA Mental Hygiene Clinic. They are being seen on the average of once every 10 days, with a mean duration of 14 months.

Group D: 18 white, male veterans who have had therapy terminated, for various reasons, in the Mental Hygiene Clinic. They were seen on the average of once every ten days. Mean duration was 7.1 months.

For the purpose of this study, the first two groups have been labeled "long-term" patients, while the last two groups have been called "short-term." Latent in Group C are patients who may be later redefined as "long-term." The first three groups are still in treatment and were tested, for the purpose of this study, while in therapy. The last group has terminated treatment.

For Groups A, B, and C, psychological testing was obtained while they were in treatment. The test battery consisted of a brief intelligence scale (Shipley-Hartford), the Minnesota Multiphasic Personality Inventory with Subtle-Obvious keys (Wiener, 1948), and Navran's Dependency Scale (1954), all scored, in addition to the traditional four validity scales and nine clinical scales with *Si* (Social introversion).

For Groups B and C, these data were also available prior to the time the veterans entered treatment. For Group D, the data were available only prior to therapy.

Table 1 represents the comparison of the groups on age, intelligence, education, and gross diagnostic category. Groups C and D are significantly younger than Groups A and B. However, there are no significant differences among the groups on education or intelligence, nor do they appear to be significantly different diagnostically.

On the MMPI, there were no significant differences between Groups A and B (long-term fee basis and clinic veterans) while in therapy, despite the fact that the former group had been in treatment a significantly longer period of time. The only scale which reflected a difference of statistical significance was *Si* (T: 2.08, $p < .05$), on which the fee-basis group (Group A) was the higher. If one does not attribute this significance to sampling error, the interpretation would be that the fee-basis group is more socially withdrawn than the group being treated in the clinic. In all other respects, Groups A and B are similar and may be thought of as constituting a single group of "interminable" patients.

In comparing Group B's pretherapy testing,

Table 1

Background of Patients Studied

	Age	Intellig.[a] (Raw Score)	Education (Years)	Gross Diagnoses Neuroses	Gross Diagnoses Pers. Disorder	Gross Diagnoses Psychoses
Group A	40.40	55.07	11.47	12	0	3
Group B	44.89	53.17	10.94	14	0	4
Group C	34.87	53.60	12.13	13	1	1
Group D	31.56	51.18	11.11	15	2	1

[a] Shipley-Hartford.

with testing after 5.3 years, there appeared to be general improvement (not statistically significant, however, and not inconsistent with the general "regression toward the mean" frequently observed in retesting with the MMPI). There was only one statistically significant reduction in scale mean over the period of therapy. That was on *Hy Obvious* (*T:* 2.07, $p < .05$), which, since it is an isolated finding and inconsistent with other data, might well have been the result of sampling error.

Group C (short-term, still in therapy) was also tested before and during treatment. After an average 14 months in therapy, there was no significant alteration on the MMPI scales. There was a tendency toward general minimal improvement, insignificant, however, with a significant drop in *D Obvious* (*T:* 2.33, $p < .05$) which, again, can be attributed to sampling error.

On pretesting, there were no significant differences between short-term terminated veterans (Group D) and short-term nonterminated (Group C), on any MMPI scales except *Hy Obvious* (*T:* 2.39, $p < .02$). When compared with long-term clinic patients (Group B), the terminated patients (Group D) also appeared not to be significantly different on any scales except *Mf* (*T:* 2.73, p .05), with the terminated patients higher.

Long-term and short-term clinic patients (Groups B and C) were also compared on pretesting. Significant differences were found on *Hs*, *Hy*, and *Hy-Obvious*, with long-term patients having the higher scores. This suggests a greater concern with physical and health problems among the long-term patients which might, in part, be accounted for by the differences in ages between the two groups. Tests of variance between all groups also yielded generally negative results.

Since the data suggested that no greater benefits were reflected in the group several years in psychotherapy, over those several months in treatment—and since the dependency of the patient (as measured by the Navran *De* scale) appeared in no measure related to the length of time spent in therapy—other possible explanatory variables needed to be explored.

To examine the influence of therapists on the duration of therapy, two samples of therapists and their patients were studied to determine the distribution of long-term patients among therapists. Tables 2 and 3 summarize this data. The first sample consisted of seven therapists and 72 patients, following therapy termination.

Using the data from Table 2, 19% of the patient population spent over 20 sessions in

Table 2

Relationship Between Therapist and Number of Interviews

Therapist	Total No. of Patients	Mn. Sessions in Therapy	No. Patients over 20 Sessions	% Patients over 20 Sessions
A	15	10.80	2	13%
B	13	40.46	7	54%
C	10	5.20	0	00%
D	18	8.89	1	06%
E	5	19.20	1	20%
F	5	17.40	2	40%
G	6	7.83	1	17%

Table 3

Relationship Between Therapist and Length of Therapy

Therapist	Total No. of Patients	Mn. Mos. in Therapy	No. over 2 Yrs.	% over 2 Yrs.	No. over 3 Yrs.	% over 3 Yrs.
A	5	25.60	1	20%	1	20%
B	3	29.67	2	67%	2	67%
C	20	12.30	2	10%	1	05%
D	13	23.61	6	46%	2	15%
E	12	17.42	1	08%	1	08%
F	16	16.69	6	38%	1	06%
G	30	14.90	2	07%	1	03%
H	31	29.61	17	55%	12	39%

psychotherapy. Using this percentage as a base line, three therapists had more of their patients in therapy over 20 sessions, and four therapists with fewer. A chi-square test of the number of long-term patients of 3 long-term therapists vs. four short-term therapists yielded a chi-square ratio of 9.69, significant at the .01 level of confidence with one degree of freedom.

A second sample consisted of eight therapists and 130 patients, all in treatment at the time of this study. This sample was somewhat more highly controlled: Only those patients who were being seen by full-time staff members with roughly equivalent training and experience were selected. Of the therapists, three were psychiatrists, two were psychologists, and three were social workers.

Table 3 summarizes the data, including the number of patients, both the mean and median months in therapy (rather than sessions this time), the percentages of patients over two years, and the percentages of patients over three years.

As can be seen from the table, there is wide variation in the percentages of patients seen for two years or longer. There does not appear to be a concentration of "long-term" patients with any particular discipline. In the first group of therapists, all three "long termers" were psychologists; in the second group, one was a psychiatrist, one a social worker, and two psychologists.

Dropping out therapists A and B (because of the low *N*s), the total number of patients seen by the remaining six therapists is 122, with a total of 2,394 therapy months and a mean of 19.6 months in therapy per patient. Total percentage of patients in treatment over two years is 27.9. Total percentage of patients in therapy over three years is 13.1. Using the latter again, as a rough cutting point, two therapists (D and H) have percentages of long-term patients greater than this. Therapists D and H account for 78% of the long-term patients in therapy. Therapist H alone accounts for 67% of all long-term patients currently being seen in the Mental Hygiene Clinic by full-time staff members.

A chi square between numbers of patients treated by the two groups of therapists yields a ratio of 15.9, significant at the .01 level of confidence, with one degree of freedom. In the second sample (the data were not readily available for the first sample), the proportions of diagnostic classes falling to each therapist do not appear to be significantly different. The percentage of diagnosed psychotics treated by the entire group of therapists was 27.7 and for the long-term therapists, 28.5. All other members of the patient group carried psychoneurotic diagnoses.

Discussion

The data presented suggest that, insofar as the MMPI and its auxiliary scales are able to assess them, long-term therapy shows few, if any, advantages over short-term therapy.

Further, the data suggest that the majority of long-term patients can be accounted for by a minority of the therapists, making it highly desirable to determine whether:

1. Certain therapists are, purposely or otherwise, assigned chronic, potentially interminable patients because of their "special talents" in dealing with them; or

2. The therapist, in his relationship with the patient, in some way encourages "interminability" because of difficulties in formulating goals, or in relinquishing relationships.

The possibility remains that there may be something inherently (and untestably) different in patients who become interminable in psychotherapy, and something that transcends rough diagnostic categories and other dimensions of behavior that are generally studied. However, it is also reasonable to examine the other half of the therapeutic relationship, as has been done previously (Auld et al., 1954; Myers et al., 1955; Myers et al., 1954; Schaffer et al., 1954) to a limited degree, to determine not only how experienced therapists differ from inexperienced therapists in relation to termination problems, but how personality types or needs within the therapists themselves affect therapy. Further investigation is needed of "interminability" qualities in therapists, since our data suggest that eclectic long-term therapy may be neither economical nor particularly beneficial.

When four judges roughly predicted who the "long-term" therapists in the study would be, their pooled rankings correlated .79 with the criterion, suggesting a predictable dimension in the area of "dependency-nurturing." The judges characterized the long-term therapists in these ways: therapist primarily desires to "help," and this feeling can be sustained indefinitely; tends to personalize relationship; tends to aim for marked changes in very sick patients; desires to succeed with intensive therapy; and needs to feel appreciated and effective (as a person and as a therapist). None of the comments made by the judges suggested that the therapists were dealing with chronic patients who required interminable support.

The results of this study point to the need for careful re-examination of therapeutic goals. Even in the event that concentration of interminable patients with a few therapists can be accounted for by "assignments of particularly chronic patients," the question of interminable treatment as a treatment of choice for these patients needs further inspection. Our tentative conclusion, that therapists themselves may be responsible for the interminability of treatment, places even greater responsibilities on the professions who prepare psychotherapists for treating chronic patients. It seems to us likely that psychotherapeutic practice today contains self-defeating concepts which may be not only hampering to the success of treatment, but potentially harmful to its clients.

Summary

1. Sixty-six patients were divided into four groups, according to the amount of time spent in therapy. Two groups were described as "long-term," and two as "short-term."

2. Follow-up psychometric findings revealed insignificant differences among the groups initially similar, suggesting that length of time in therapy does not correlate with improvement.

3. Two samples of therapists indicated that the great majority of long-term patients were being seen by a small minority of the therapists. A likely reason appears to be failure on the part of "long-term" therapists to adequately formulate therapy goals, and to impersonalize the therapeutic relationship.

Received May 2, 1958.

REFERENCES

Auld, F., & Myers, J. K. Contributions to a theory for selecting psychotherapy patients. *J. clin. Psychol.,* 1954, **10**, 56–60.

Conrad, D. C. The duration of the therapeutic relationship and therapists' successive judgments of patients' mental health. *J. clin. Psychol.,* 1954, **10**, 229–233.

Federn, P. *Ego psychology and the psychoses.* New York: Basic Books, 1953.

Freud, S. Analysis terminable and interminable. In *Collected papers.* Vol. V. London: Hogarth, 1953. Pp. 316–357.

Imber, S., Nash, E., & Stone, A. Social class and duration of psychotherapy. *J. clin. Psychol.,* 1955, **11**, 281–284.

Kotkov, B., & Meadow, A. Rorschach criteria for predicting continuation in individual psychotherapy. *J. consult. Psychol.,* 1953, **17**, 16–20.

Masserman, J. H. *Practice of dynamic psychiatry.* Philadelphia: Saunders, 1955.

Myers, J. K., & Auld, F. R. Some variables related to outcome of psychotherapy. *J. clin. Psychol.,* 1955, **11**, 51–54.

Myers, J. K., & Schaffer, L. Social stratification and psychiatric practice: A study of an outpatient clinic. *Amer. sociol. Rev.,* 1954, **19**, 307–310.

NAVRAN, L. A rationally derived MMPI scale to measure dependence. *J. consult. Psychol.*, 1954, **18**, 192.

RANK, O. *The trauma of birth.* New York: Harcourt, Brace, 1929.

ROGERS, C., & DYMOND, R. *Psychotherapy and personality change.* Chicago: Univer. Chicago Press, 1954.

ROGERS, L. S., KNAUSS, J., & HAMMOND, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.*, 1951, **15**, 45–57.

SCHAFFER, L., & MYERS, J. K. Psychotherapy and social stratification: An empirical study of practice in a psychiatric outpatient clinic. *Psychiatry*, 1954, **17**, 83–93.

WEIGERT, E. Contribution to the problem of terminating psychoanalyses. *Psychoan. Quart.*, 1952, **21**, 465–480.

WIENER, D. N. Subtle and Obvious Keys for the MMPI. *J. consult. Psychol.*, 1948, **12**, 164–170.

WINDER, A., & HERSKO, M. The effect of social class on the length and type of psychotherapy in a VA Mental Hygiene Clinic. *J. clin. Psychol.*, 1955, **11**, 77–79.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

GSRs DURING REPRESSION, SUPPRESSION, AND VERBALIZATION IN PSYCHOTHERAPEUTIC INTERVIEWS

JESSE E. GORDON[1]
Montana State University

BARCLAY MARTIN AND RICHARD M. LUNDY
University of Wisconsin

The relationship between anxiety and the process whereby a psychotherapy patient becomes aware of and eventually verbalizes repressed material is a matter of both theoretical and practical interest to the therapist. The present status of psychological theory with respect to this problem allows for several different and contradictory predictions about this relationship.

It is perhaps reasonable to assume that when certain material is repressed, the repressor experiences anxiety associated with this material to the extent that the repression is incomplete. That is, with complete repression, internal anxiety-producing cues are not present and, accordingly, an anxiety response is not made. As repression becomes less complete (more and more anxiety-producing cues appear), more and stronger anxiety responses are made. From this point on, it is possible to pose reasonable but alternative predictions as to the relationship between anxiety and the process of "lifting" a repression.

On the one hand is the view implied in some psychoanalytic writings that the achievement of full realization ("insight") into the repressed material results in relief or anxiety reduction.

On the other hand, one might expect that the increased number and clarity of the anxiety-producing cues, associated with less and less complete repression, leads to even greater anxiety as the individual approaches complete awareness of the formerly repressed material. Further, overt verbalization of the repressed material should involve even more fully constituted anxiety-producing cues and, accordingly, greater anxiety than is present during awareness without overt verbalization.

To complicate the problem further, one must introduce the possibility of extinction of the learned association between cue and anxiety response. Such extinction is likely to begin shortly after awareness or overt verbalization has occurred in psychotherapy.

In view of the above considerations it does not seem appropriate to state directional hypotheses in the present study, which was designed to explore empirically the relationship between anxiety and the "lifting" of a repression, as the threatening material moves from a state of repression through awareness without verbalization (suppression) to overt verbalization.

Research which might clarify this relationship has been made difficult by the nature of repression. An experimenter cannot know what is repressed until it is no longer repressed, at which time it is too late to measure anxiety associated with it. However, a set of procedures used in an earlier study (Gordon, 1957) has been modified to make possible measurement of anxiety in a situation in which repression and suppression of known content is experimentally induced. Thus measurement of anxiety during repression, suppression, and verbalization of experimenter-controlled material is possible.

A state of repression is assumed to exist when a person fails to produce either overt or covert cues (usually verbal) to which anxiety responses have been attached, where the production of such cues is appropriate. For the purposes of this research, posthypnotic instructions were used for inducing this condition. Posthypnotic amnesia was created by the instruction to "not think about" a particular area of experience. The area of experience selected was that of conflict with one or both parents. It was assumed that anxiety is usually associated with such experiences, so that the words used to symbolize such experience also serve as cues for anxiety.[2]

---

[1] This research was performed while J. E. Gordon was Resident Lecturer in Psychology at the University of Wisconsin, supported by a U.S.P.H.S. grant.

Suppression is defined as an awareness of an area of experience together with the motivation to not verbalize about it overtly. The experimental manipulations involved hypnotic instructions that the subject would think about and recall her conflicts with her parents, but would not want to talk about them.

Verbalization was defined as conversation about parents and/or conflicts with parents with the interviewer. The experimental manipulations involved hypnotic instructions to want to talk about conflicts with parents with the interviewer.

Anxiety was defined in terms of galvanic skin conductance. Although there is disagreement about the nature of anxiety and the most valid measures of it, there is considerable evidence that skin conductance, which reflects changes in sympathetic nervous system activation, is associated with states of general emotional arousal or mobilization (Schlosberg, 1954). Emotional arousal may be a somewhat broader concept than anxiety, but lacking precise operational distinction between anxiety and this more general concept, the skin conductance measure was accepted as the operational definition of anxiety in this study.

Procedure

Subjects. Ss were 12 female undergraduate students at the University of Wisconsin. These Ss were selected from a group of 20 females who had already been trained and screened for hypnotic susceptibility, who agreed to serve in this research. The group of 20 from which they came was screened from an initial group of 44 females who volunteered for "an experiment on hypnosis." Training and screening procedures consisted of three successive hypnotic sessions, in each of which a more "difficult" posthypnotic suggestion was given, and the response tested. The final criterion for selection of Ss was complete posthypnotic amnesia by the end of the third training and screening session.

Control. After the third session, appointments were made with Ss for a control interview. Immediately preceding this interview, S was told that electrodes would be attached to her fingers in order to record "changes in the electrical activity" of the skin, and S was reassured that she would experience no electrical shock. S was then introduced to the interviewer who attached the electrodes to S seated in a chair at the interviewer's desk. Control interviews consisted of 15-minute nondirective interviews conducted by a trained clinical psychologist. Ss were told that they could talk about anything they cared to, that the topic of conversation was completely up to them. The purpose of this interview was to provide a period of adaptation to the GSR device, and to obtain a measure of base level conductance against which to compare the conductance measure obtained under the experimental conditions.

At the conclusion of the control interview, each S was again seen by the principal investigator, who asked her to fill out the Psychotherapy Reaction Questionnaire, a scale developed by Ford (Ashby, Ford, Guerney,

---

[2] One may question these procedures as satisfying the definition of Freud's concept of repression as a defensive maneuver. The instructions to "not think about," while omitting specific mention of avoidance motivation, however, do presumably produce a phenomenon in which previously "conscious" material is no longer verbalizable (i.e., becomes "unconscious") even though this material later becomes available for verbalization. Thus we are not dealing with forgetting or extinction. The process in the subject by which material is for a time rendered unconscious is better described by the term "repression" than any other term in current usage. As far as defensive motivation is concerned, it is assumed that because the conflict material used in this study is usually unpleasant to most people, then anxiety is assumed to be associated with the verbal conflict symbols. There is reason to regard such associated anxiety as a drive or motivational factor, the reduction of which can reinforce the responses which led to the anxiety reduction. In this sense, the instructions to "not think about" an unpleasant experience are considered to be analogous to defensively motivated repression.

& Snyder, 1957) to tap defensive reactions and positive or negative attitudes toward the therapist.

Experimental conditions. Two experimental groups were formed. The only difference between these groups was in the order of the experimental conditions. One group, designated RVS, was to receive instructions designed to produce repression, verbalization, and suppression in that order. The other group, designated RSV, was to be given instructions designed to produce repression, suppression, and verbalization in that order.

Following the control interviews, *S*s were appointed in random order to the two groups for the fifth and final sessions. *S*s were hypnotized and given the following instructions:

> Now, as you continue to sleep, I want you to recall in your imagination some of the occasions on which you have come into conflict with either or both of your parents. Think of times when you might have had arguments, or when you felt either your mother or your father were too strict, or when you might have felt that you were unfairly treated or not being understood. Review some of these occasions in your memory. See them in your mind's eye. You don't need to tell me about them—just think about them yourself. I'll give you a minute or two in which I'll be silent to give you a chance to envision yourself in those situations. Feel just exactly the way you felt at those times. Kind of relive those experiences in your imagination now. Just think about them for a little while.

Conflict with parents was selected as the area of experience for this research because it is a fairly universal experience, and thus likely to be realistic for each *S*, and is one which is usually attended by anxiety. The procedures thus have the advantage over artificially induced neuroses of permitting some control over a true and meaningful experience in *S*'s past.

Following these instructions, *S*s in the RVS group were instructed as follows:

> Now, when you wake up, all those experiences you've been thinking about will have completely left your mind. You won't be thinking of them at all. They'll be the farthest things from your mind. In the beginning of your interview with Dr. ——, he'll click a key three times. Just ignore that. It doesn't mean anything. But later on, when he clicks the key again, all these thoughts and experiences with your father or mother or both, and your feelings about those experiences, will come to mind. You'll begin to think about them again. You can talk about the feelings and experiences that are on your mind. You'll want to talk them over with Dr. ——. Later on, Dr. —— will click the key again. When he does, it'll occur to you that you shouldn't really talk to him about your experiences and feelings toward your parents, and you won't want to talk about them anymore, although they will still be on your mind.

These instructions were then summarized, posthypnotic amnesia for the instructions induced, and *S* awakened.

For the RSV group, the instructions were substantially the same, except that instructions for suppression were given before the instructions for verbalization.

Upon awakening, *S* was told that she was to be interviewed by Dr. ——, and that throughout the course of the 45-minute interview, she would have electrodes strapped to two fingers. *S* was told that, as in the control interview, she could talk about anything she cared to. *S* was then brought to the interview room, introduced to Dr. ——, and the electrodes attached. The interviews began in a standard manner, with the interviewer telling *S* that the interview would be about the same as the one she had before and that she might start off by telling something about herself. From that point on, the interview remained within the nondirective framework. All interviews were tape recorded. At the conclusion of the interview, *S* returned to the experimenter who again administered the Psychotherapy Reaction Questionnaire, disabused *S* of the hypnotic instructions, and told her of the purpose of the study.

The interviewer did not know beforehand whether an *S* was in the RVS or RSV group.

GSR recording. Skin conductance measures were obtained by using an AC circuit designed by Grant (1946) and dry silver electrodes bound to the palmar surfaces of two fingers of *S*'s left hand. Continuous ink recordings of *S*'s resistance throughout the session were obtained by feeding the output of the GSR apparatus into a General Electric Photoelectric Recorder. Resistance scores were obtained by halving the difference between the low and high points in one-minute sections of the ink recordings, and adding this to the low point within each one-minute interval. The conductance scores for data analy-

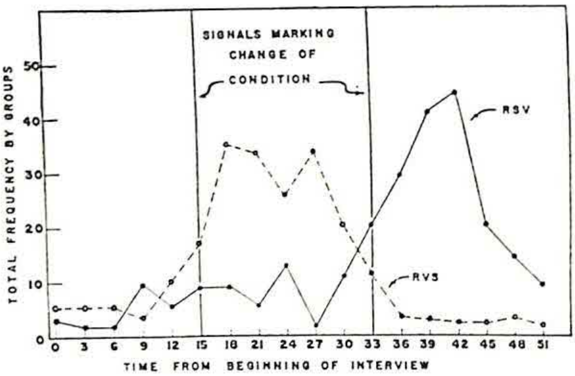


Fig. 1. Distribution of references to parents and conflict with parents.

sis were the reciprocals of these resistance measures.

Results

As a check on the validity of the experimental procedures, the interview tape recordings were scored in 30-second intervals for the presence of reference to parents, and to conflict with parents. Each interval was scored plus or minus for the presence or absence of such references during that interval. No attempt was made to score for magnitude of references, or to frequency within intervals. Figure 1 shows the distribution of plus scores for the RVS and RSV groups. It is apparent from Fig. 1 that the *S*s did follow the hypnotic instructions, as far as verbalizing is concerned.[3] One *S* in each group failed to follow the instructions; neither made any reference to parents, with or without conflict, throughout the course of her interview. These two *S*s were therefore dropped from further analysis.

Figure 2 presents the conductance scores of both groups over all conditions.[4] The abscissa represents number of two-minute intervals from the beginning of the particular experimental condition. As noted above, the GSR records were scored in one-minute intervals. These data were collapsed into two-minute intervals by adding the scores together in successive pairs of one-minute intervals. The height of the curves therefore represents the RVS and RSV means of conductance scores (in micromhos) summed over two one-minute intervals.

It may be seen from Fig. 2 that there is a general increase in conductance across the entire time period. The levelling-off of the conductance, usually interpreted as adaptation, which is commonly found in GSR records after the first 15 minutes of recording, is not present in these data. That the over-all increase in height is not a function of the particular order of the experimental conditions is suggested by the data on Suppression and Verbalization. The conductance curve for the last experimental condition is higher than for the next to last, regardless of whether the last was Suppression or Verbalization.

The GSR data were therefore analyzed for differences in shape of the conductance curves, rather than for differences in mean conductance levels. It is thus possible to draw conclusions about the relationship between the independent variables and the growth of anxiety, although level of anxiety at any particular point must be ignored. Figure 3 pre-

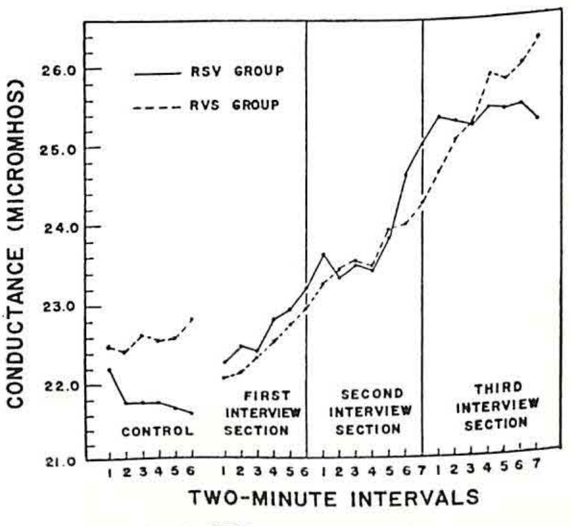


Fig. 2. GSR curves over time.

---

[3] The lines marking the boundaries of the experimental conditions are approximate. The interviewer varied somewhat in his timing of the signals, so as not to interrupt an *S* in a flow of conversation. The interview sections are therefore not strictly comparable from *S* to *S* in terms of length. The boundary lines in this figure are drawn at the times when the signals were most frequently given.

[4] As was noted in connection with Fig. 1, the length of the experimental conditions varied somewhat from *S* to *S*. The number of intervals reported in this figure and used in the statistical analyses was the maximum number of intervals common to all *S*s. For several *S*s, therefore, the last one or two intervals in some conditions were dropped from the analyses.

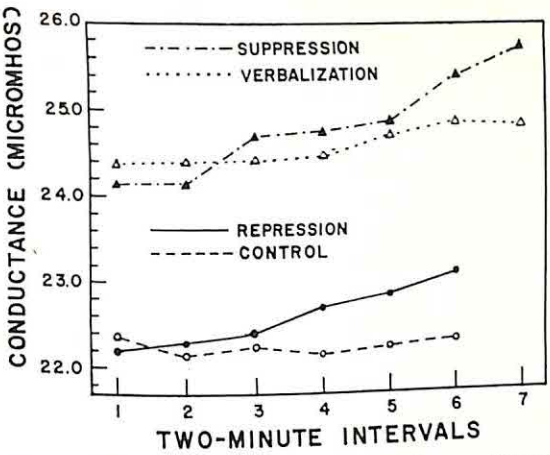


Fig. 3. GSR curves of all *S*s combined, by conditions.

sents the conductance curves of all *S*s, combined by experimental conditions. The units are the same as in Fig. 2.

Grant's (1956) extension of Alexander's trend test was applied to these data. For each *S*, a difference was found between her conductance levels in two-minute intervals of one condition and the comparable two-minute intervals in a comparison condition. Thus a set of difference curves was generated for each pair of conditions which were to be compared with each other. Each difference curve was tested for the significance of its difference from a curve of zero slope. This then tells one whether there is a difference in the slopes of the curves representing the different conditions to be compared. Tables 1, 2, 3, and 4 report the results of these analyses.

Table 1 indicates that under the Repression condition, *S*s' conductances were increasing at a faster rate than they were during the Control interviews. It appears that anxiety increases as *S* approaches the point of awareness, and the repression mechanism becomes less effective.

Tables 2 and 3 compare Repression with Suppression and Verbalization. The results do not reveal any significant differences between the conductance curves in Repression and Suppression. The Repression-Verbalization differ-

Table 1

Analysis of Variance of Trend of Differences Between Repression and Control[a]

Source	*df*	Mean Square	*F*
Between ind. slopes	8	1.46	
Between ind. quadratics	8	.18	
Between ind. cubics	8	.28	
Over-all slope	1	6.96	4.77*
Over-all quadratic	1	.10	.56
Over-all cubic	1	.12	.43

* $p < .05$.

[a] Owing to error in recording, control GSR level of one *S* was useless for statistical analysis. Therefore this analysis is based on 9 *S*s.

Table 2

Analysis of Variance of Trend of Differences Between Repression and Suppression

Source	*df*	Mean Square	*F*
Between ind. slopes	9	.86	
Between ind. quadratics	9	1.13	
Between ind. cubics	9	.06	
Over-all slope	1	.67	.78
Over-all quadratic	1	.24	.21
Over-all cubic	1	.02	.33

Table 3

Analysis of Variance of Trend of Differences Between Repression and Verbalization

Source	*df*	Mean Square	*F*
Between ind. slopes	9	1.63	
Between ind. quadratics	9	.32	
Between ind. cubics	9	.89	
Over-all slope	1	1.78	1.09
Over-all quadratic	1	.001	—
Over-all cubic	1	.001	—

Table 4

Analysis of Variance of Trend of Differences Between Suppression and Verbalization

Source	*df*	Mean Square	*F*
Between ind. slopes	9	1.23	
Between ind. quadratics	9	1.79	
Between ind. cubics	9	.15	
Over-all slope	1	7.20	5.85*
Over-all quadratic	1	.27	—
Over-all cubic	1	.20	1.33

* $p < .05$.

ence curve also does not yield a significant slope, although there is a tendency for the Repression curve to be somewhat steeper than the Verbalization curve.

Table 4 reports the comparison of Suppression and Verbalization. The significant difference found indicates that conductance measures obtained under Suppression increase at a greater rate than those obtained during Verbalization. Figure 3 reveals that the Verbalization slope is almost flat, whereas the Suppression slope is one of the steeper slopes obtained. Thus it is found that anxiety accumulates or builds up at a faster rate during Suppression than during Verbalization of disturbing material.

The differences between the RSV and RVS groups in the questionnaire of attitudes toward the therapist yielded $t = 1.64$ ($p > .05$). The mean change in the RSV group was +3.6, which is in the direction of more positive attitudes; the mean change in the RVS group was -5, which is in the direction of more negative attitudes; $\sigma_{\bar{x}-\bar{y}} = 5.25$. Although the differences are not significant, due to the great variability among the *S*s, the results suggest a tendency for RSV *S*s to feel more positively toward the therapist than the RVS *S*s. This suggests that *S*s whose most recent experience has been one in which they revealed, rather than hid from the therapist, some personally threatening material may feel more positively toward the therapist.

Discussion

The most striking outcome of this study is the finding that there is never a downward turn in the absolute measure of anxiety during the 45-minute experimental session. This suggests that the lifting of a repression, whether it included Suppression or Verbalization, or both, is accompanied by a fairly consistent rise in anxiety. An absolute reduction in anxiety, for whatever reason, does not seem to occur during the first 30 minutes or so after the *S* first becomes aware of the repressed material. It is of interest to note that the two *S*s who failed to comply with the posthypnotic instructions by never verbalizing about parents in any fashion were the only two *S*s who did not show this over-all increase in conductance. One of these *S*s had a decreasing conductance curve, and the curve of the other *S* was about level. Experience with GSR indicates that this upward trend is not necessarily characteristic of interviews of comparable length. This phenomenon thus appears to be associated with the effects of the hypnotic instructions. One possible explanation of the continuing rise in conductance may be that anxiety associated with parental conflict accumulates throughout the time that the *S*s are forced by the hypnotic instructions to deal with this aspect of their experience. The over-all upward trend does tend to level off during Verbalization. One may account for this by assuming that extinction of the association between cue and anxiety response occurs when anxiety-producing material is verbalized in an accepting atmosphere. From this point of view, we might expect the curve eventually to turn downward if sufficiently long periods of Verbalization were maintained until the verbal cue-anxiety association is extinguished. Perhaps the 15-minute period was not sufficient to get a full indication of the degree to which extinction could occur during Verbalization.

A simpler account of the change in slope during Verbalization might be that the increase in anxiety is retarded only when the *S* is talking about the anxiety-arousing material. In this sense, the anxiety curve begins to level off only when there is congruence between the *S*s' experiencing of the material, their awareness of the material, and their verbalization of it. This account also suggests that, had there been more time in which *S* could experience this congruence, the anxiety might have reduced. When there is not this complete congruence, as in the Repression and Suppression conditions, anxiety consistently increases.

Given this over-all increase in conductance, which appears to be unrelated to the particular experimental conditions, except possibly for Verbalization, one may still compare the conditions in order to discover which ones contribute most heavily to the growth of the anxiety. The results of such comparisons indicate that repressing *S*s generate anxiety at a faster rate than they do in the nonrepressing control condition. One cannot decide from these data whether the increasing slope in repression is caused by the imminent cessation of repression, or by a simple anticipation of the signal denoting the start of another condition. The control interview was not followed by another condition, as was the case with the Repression section. However, an examination of the curves in Fig. 2 suggests that increases in conductance may not be solely determined by anticipation; the RVS group does not have as rapid an increase in conductance as the RSV group during the second interview section, even though there is still a third interview section for both groups to anticipate. It is suggested that what is being anticipated plays at least some part in the anticipation response. This further suggests that some anxiety-arousing cues are being produced in the *S*s even though they are repressing, as defined by the hypnotic instructions, and that these cues increase as the point of awareness approaches. Since the *S*s are not overtly verbalizing about the induced threat or about the signal which they may be anticipating, it must be assumed that the anxiety-arousing cues are being produced covertly. If we assume that the *S*s were not thinking about the threat or the approaching signal, according to the hypnotic instructions, we must conclude that the anxiety-producing cues were being produced without the *S*s' awareness.

This raises some questions about the nature of awareness. According to Dollard and Miller (1950), repression exists when an individual fails to think (produce verbal cues, usually) because of the association between these cues and anxiety. This is essentially a learned avoidance response. If such were the case, one would expect little or no anxiety in repressing *S*s. This, of course, is not what was found. Our data, indicating increasing anxiety during Repression, suggest that some anxiety-arousing cues were being produced when the *S*s are assumed, on the basis of the hypnotic instructions, to be unaware of the threat. Thus there is evidence for the presence of anxiety-producing cues without awareness. Dollard and Miller's hypothesis that unawareness or unconsciousness is defined in terms of the absence of such cues is therefore inconsistent with the analysis based on our data. It is suggested that awareness be defined in terms of some other secondary response to the covert (verbal) anxiety cue-producing responses which the unaware *S*s appear to be making, and that it is these other responses (perhaps self-related evaluations) which are not being made by repressing *S*s. That is, awareness may be defined in terms of some self-related evaluations which the individual may or may not make as responses to internal, usually verbal cue-producing responses, rather than in terms of the cue-producing responses themselves.

Hypnosis may be thought of as an indirect way of producing the condition of not being aware of the covert cues associated with anxiety. There is no guarantee that the *S*s were indeed unaware of the anxiety-producing cues, namely, thoughts about parental conflict. If one is willing to accept the validity of the posthypnotic instructions in this respect, then the rather interesting conclusion must be drawn that anxiety increases with equal rapidity whether *S* is aware of the covert anxiety-producing cues or not, since the conductance curves for Repression and Suppression do not differ in slope. This finding that anxiety continues to rise in the Suppression condition suggests that awareness or insight, of itself, is not sufficient to reduce anxiety.

At the same time, an account of the relationship between anxiety and the lifting of repression which would predict a direct relationship between anxiety and availability of anxiety-producing cues would have some difficulty with these data. With initial verbalization and its profusion and clarity of cues, we should expect an even greater increase in the anxiety measure. In contrast we find that it is only during Verbalization that the increase in anxiety tends to slow down.

This reduction in anxiety increase during Verbalization may have a counterpart in the positive feelings toward the therapist which the RSV *S*s tend to show to a greater extent than the RVS *S*s. It seems probable that the verbalization section, occurring last for the RSV *S*s, leaves the *S* with a more positive feeling because the *S* has been experiencing a lessening of the anxiety increase just prior to her rating of the therapist. In contrast the increase in anxiety associated with the last condition—Suppression—for the RVS *S*s may be associated with the tendency of these *S*s to feel less positive toward the therapist.

Certain qualitative aspects of the *S*s' behavior in the interviews may provide a clearer picture of the nature of these interview sessions. In all respects, *S*s' verbal behavior was natural; it is doubtful that the most insightful psychologist would discover, from listening to the tape recordings, that *S*s were performing according to hypnotic suggestions. Of particular interest is the manner in which *S*s responded to the signal denoting the beginning of a new experimental condition. In no case was there a verbal response relating to the signal itself. Some *S*s lapsed into silence for a few seconds, and then began verbalizing, usually with a change to subject matter appropriate to the new condition. A frequent remark from *S*s who had been in the Verbalization condition when the signal was given marking the beginning of Suppression was "Well, let's talk about something pleasanter," or "I guess I shouldn't bore you with these personal problems." During Suppression sections, particularly in the RSV group, one notes frequent references to the parental conflicts of others—friends, relatives, etc., and occasionally a spontaneous denial of personal conflicts with parents.

One recording was particularly illuminating in the manner in which it reveals the development of associations to the threatening material. The *S* was in the RSV group. The major topic of conversation during Repression was *S*'s interest in a film on intellectual deficiency and brain pathology she had recently seen in psychology lecture. This film reminded her of her friend's mother, who suffers from multiple sclerosis. This conversation was continued in the Suppression section, where the major emphasis was on the emotional climate of the friend's home, in relation to the mother's illness. *S* stressed the frequent conflicts between her friend and her friend's father, growing out of the continual stress at home. This continued until the Verbalization section began, in which *S* discussed her own conflicts with her parents. She introduced the topic by saying, "Of course, lots of people have problems with their parents, even when there isn't that kind of sickness and tension." Thus the discussion of the psychology film seems to have been closely associated with, or at least led directly to, a discussion of *S*'s own problems. This may suggest that therapists be cautious in deciding that any particular topic is as irrelevant as a discussion of a film on brain damage at first appears. It is the ability to produce behaviors such as these which become intelligible only when the experimenter has knowledge and control of the antecedent conditions, that makes the experimental procedures used in this research of particular value.

Summary

An experiment was performed to study the effects on anxiety, as measured by GSR recording, of repression, suppression, and verbalization of threatening material. Ten selected and trained undergraduate female *S*s were stimulated to recall conflicts with their parents under hypnosis, and were given posthypnotic suggestions to "not think about" (Repression), "think about but not talk about" (Suppression), and "talk about" (Verbalization) conflicts with parents to the interviewer, who conducted a nondirective type interview for 45 minutes with each *S*. Each condition occupied 15 minutes of the interview, with *S* responding to signals demarcating the beginning of each condition as instructed under hypnosis. Skin conductance was recorded continuously throughout the interviews, and a scale of attitudes toward the therapist given once at the end of a 15-minute control interview and again at the end of the 45-minute interview, constituted the dependent variables. Skin conductance measures were analyzed in terms of differences in slope. The findings were as follows:

1. There was a consistent over-all increase in conductance throughout the experimental session, indicating a continuous growth in anxiety regardless of the order of conditions.
2. The increase in conductance was more rapid under Repression than during the control interview.
3. Conductance slopes during Repression and Suppression were not found to differ significantly.
4. The conductance slopes during Suppression were significantly steeper than during

Verbalization, in which the slopes tended to flatten.

5. *S*s whose last interview section was Verbalization showed a tendency toward more positive shifts in attitudes toward the therapist than *S*s whose last interview section was Suppression. However, this difference was not significant.

Received May 9, 1958.

REFERENCES

Ashby, J. D., Ford, D. H., Guerney, B. G., Guerney, Louise F., & Snyder, W. U. Effects on clients of reflective and a leading type of psychotherapy. *Psychol. Monogr.*, 1957, **71**, No. 24 (Whole No. 453).

Dollard, J., & Miller, N. *Personality and psychotherapy.* New York: McGraw-Hill, 1950. Pp. 198–221.

Gordon, J. E. Leading and following psychotherapeutic techniques with hypnotically induced repression and hostility. *J. abnorm. soc. Psychol.*, 1957, **54**, 405–410.

Grant, D. A. A convenient alternating current circuit for measuring GSR's. *Amer. J. Psychol.*, 1946, **59**, 149–151.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Schlosberg, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

A NOTE ON REMOTENESS OF TAT FIGURES AS AN INTERPRETIVE CONCEPT[1]

EVERETT PILE, WALTER MISCHEL[2]
University of Colorado

AND LEWIS BERNSTEIN
Veterans Administration Hospital, Denver, Colorado

Tomkins (1947) has introduced the concept of remoteness between the TAT narrator and the TAT figures as an important interpretive device. This concept assumes that a hierarchy of impulse acceptance can be established for the narrator. For example, Piotrowski (1950) has stated: ". . . the most acceptable drives are projected into TAT figures who resemble the testee by age, sex, race, social status, etc. Less acceptable drives are ascribed to figures of the same sex but of different age. Actions which are still less acceptable are projected into figures of the opposite sex. Tendencies that possess a still higher degree of unacceptability are assigned to figures both of opposite sex and of different age" (p. 108). Van Lennep (1948, 1951) also has written extensively on this general theme and regards the individual who utilizes projection as manifesting an attempt to "distance" himself from ego-threatening self-attributes.

The present study was designed to test the validity of Piotrowski's assumption, and the following hypothesis was formulated for this purpose: The less acceptable certain behavior is to the narrator, the more he will attribute such behavior to TAT figures dissimilar to himself along a continuum of age and sex.

Method

Twenty-one male neuropsychiatric patients at the Denver Veterans Administration Hospital were used as *S*s. They were selected from patients referred to the Psychology Service for diagnostic testing. No psychotic patients were used since it was felt that there would be a higher probability of their misunderstanding the task.

*S*s were asked to write TAT stories using Cards 1, 4, 6GF, 7BM, 7GF, and 13F. These cards were selected since they include approximately an equal number of males and females (five and six, respectively). Further, the figures on these cards represent a wide age range. The female figure in the background of Card 4 was not used, as only two *S*s referred to it.

*S*s wrote the stories with the experimenter absent from the testing room. This procedure has been found by Bernstein (1956) to elicit freer expression of feeling in TAT stories. The following instructions were read to each *S* and left with him in written form, after it was felt that he understood the nature of the task:

> This is a test of imagination, one form of intelligence. The cards which you have been given have pictures on them. You are to look at them, one at a time, and write as dramatic a story as you can for each one. Tell what has led up to the event shown in the picture, describe what is happening at the moment, what the characters are feeling and thinking, and then give the outcome. Try to spend about five minutes on each story.
>
> When you have finished writing the story, guess the age of each of the people in the picture and write the ages at the bottom of the paper upon which you have written the story.

To test the hypothesis, it was necessary to quantify the variables of (*a*) acceptability-of-behavior for each TAT figure, and (*b*) the age-sex dissimilarity between *S* and each TAT figure.

---

[1] The authors wish to express their appreciation to Robert Fadeley, John Thompson, and Virgil Willis, who assisted in testing some of the subjects.

[2] Now at Harvard University.

The acceptability-of-behavior of each TAT figure was measured by first classifying all the scoreable behavior in *S*'s stories according to the scheme of needs proposed by Murray (1938). For example, if *S* wrote about one figure, "This man has just shouted angrily at the other person," this behavior was classified under the need, Verbal Aggression. After all the needs had been extracted from *S*'s stories, he was seen in a second session. He was given a pack of small cards, each of which contained a behavioral description representing one of the needs he had expressed in some story,[3] and asked (under the guise of this being another test) to rank-order these behavioral descriptions in terms of their acceptability to him. Instructions were:

> In this test, read all the cards carefully. Then arrange them on the table in a column according to how acceptable the behavior described on each is, from your point of view.
>
> When you finish, the card with the behavior you feel is most acceptable will be at the top of the column and the card with the least acceptable behavior will be at the bottom. The rest of the cards in the column will be in order of their acceptability.
>
> Think of "acceptable behavior" as the kind which you would be comfortable in claiming as your own, the kind you feel is the best to use, generally speaking.
>
> Take your time and feel free to move the cards around until you have them in just the order that you want them.

The rank numbers assigned to the needs *S* had expressed were used to compute an acceptability-of-behavior score for each TAT figure. This score was found by summing the rank numbers of all the needs attributed to a particular figure. For example, if a figure's behavior was described by three needs, the rank numbers given these needs were summed, and the sum became the acceptability-of-behavior score for that figure. These scores for the eleven TAT figures were then rank-ordered, thus arranging the figures in a hierarchy of acceptability-of-behavior.

The second variable, dissimilarity between *S* and TAT figures, was quantified by ordering the figures onto a continuum of age and sex. Differences were computed between the age of *S* and the age he estimated for each TAT figure. The rank-order differences for the five male figures were assigned Ranks 1–5, and the age differences of the female figures provided Ranks 6–11. By this method, both age and sex differences were included in the dissimilarity continuum, as required by the hypothesis.

Finally, a rank-order correlation between the ranks of the acceptability-of-behavior scores and the dissimilarity ranks of the 11 TAT figures was computed for each subject.

Results and Discussion

Of the 21 *S*s, 10 wrote stories in which needs were attributed to each of the TAT figures. A rho correlation for each of these 10 *S*s was computed between the ranks of the acceptability-of-behavior scores and age-sex dissimilarity ranks of the total 11 TAT figures. This group is referred to as "nondistanciators." It will be noted that none of these correlations, shown in Table 1, are significant.

A rho correlation was computed for each of the remaining *S*s who did not attribute needs to all the figures in the stories written.

Table 1

Correlations Between Acceptability-of-Behavior and Age-Sex Dissimilarity Scores

*S*s Attributing Needs to All TAT Figures		*S*s Failing to Attribute Needs to All TAT Figures	
S	Rho	*S*	Rho
1	−.29	4	.75*
2	−.31	5	.96**
3	−.26	6	.62*
7	−.21	8	.82*
9	.28	12	.69*
10	−.07	14	.73*
11	.45	15	.81**
13	.24	16	.75*
17	−.11	18	.68*
20	−.11	19	.51
		21	.85**

* Significant at .05 level.
** Significant at .01 level.

---

[3] For example, the description for the need, Verbal Aggression, was: "He makes fun of people, is sarcastic, and quarrelsome."

The correlation for each of these *S*s is between the ranks of the acceptability-of-behavior scores and age-sex dissimilarity ranks of the TAT figures to which the particular *S* had attributed needs. Thus, the rhos from these 11 *S*s are based upon the data from different numbers of TAT figures which vary from 5 to 10. These individual correlations also are shown in Table 1. Of these 11, 10 are significant at or beyond the .05 level of confidence. This group of Ss is referred to as the "distanciators."

These findings focused attention upon the individual differences in distanciation, i.e., the fact that *S*s who failed to attribute needs to all the TAT figures did distanciate, whereas *S*s who attributed needs to all the figures did not. This raises the question of what subject variables may be producing such variance. In line with this, the length of hospitalization for the two groups was examined. It was found that the distanciators had been hospitalized for a mean period of 4.40 weeks while the nondistanciators had been hospitalized for a mean period of 8.64 weeks. The difference in length of hospitalization between the two groups, using the Mann-Whitney *U* test, yields a *U* of 24 which, for the size of the two groups, is significant at the .025 level. (The sigmas of the distanciators and nondistanciators are, respectively, 4.23 and 7.51.)

In retrospect, this finding is logically consistent with some emerging theoretical formulations which the authors are beginning to construct regarding the manner in which unacceptable behavior is handled. It would appear that the degree of distanciation manifested is related to certain factors which are part of the situation in which *S* behaves. For example, the *S*s in this study represent hospitalized, emotionally disturbed individuals, most of whom were currently having difficulty with their usual defenses of repression and denial. For some, it was the first time in a therapeutic environment [4] which typically encourages a frank examination of behavior. This, plus the experience of psychotherapy which most were receiving in some form, might operate to bring their needs closer to awareness. This possible factor, coupled with the reduction in fear of realizing and admitting the existence of many undesirable personal characteristics could vary with the length of time in the situation (the therapeutic environment). This possibility is consistent with the finding that the distanciators, hospitalized for a shorter length of time, failed to attribute needs to *every* TAT figure. Their stories are thus more evasive, if we interpret this failure to attribute needs to all the figures in TAT stories as evasion.

The results seem to indicate that the assumption set forth by Piotrowski holds, but only under certain conditions. The next task must be to study more fully the factors in the situation which comprise these conditions. Further research is now in progress to determine whether differential stress experimentally generated by verbal cues affects the degree of distanciation manifested in TAT stories.

Summary

This study tested an assumption of Piotrowski that the less acceptable certain behavior is to the TAT narrator, the more he will attribute such behavior to TAT figures dissimilar to himself along a continuum of age and sex. Twenty-one male neurotic patients at the Denver VA Hospital wrote stories to six selected TAT cards, and estimated the age of each TAT figure. These stories were scored for a set of standard needs. During a second session, each *S* rank-ordered all the needs which he had attributed to the TAT figures, using a criterion of "personally acceptable behavior."

A rho correlation for each *S* was computed between the ranks of the acceptability-of-behavior scores and the age-sex dissimilarity ranks. Ten *S*s wrote stories attributing needs to all the TAT figures; none of the correlations for these Ss was significant. Eleven *S*s failed to attribute needs to all TAT figures; 10 of the 11 correlations for these *S*s were significant.

These findings demonstrate that Piotrowski's assumption is valid only under certain conditions. In the present study, length of hospitalization appears to be a relevant variable.

---

[4] Since the hospital in which this study was conducted is a teaching center, only patients with acute psychiatric conditions which are expected to respond to relatively short-term intensive treatment are admitted.

Received May 12, 1958.

REFERENCES

Bernstein, L. The examiner as an inhibiting factor in clinical testing. *J. consult. Psychol.*, 1956, **20**, 287–290.

Murray, H. A. *Explorations in personality.* New York: Oxford, 1938.

Piotrowski, Z. A new evaluation of the Thematic Apperception Test. *Psychoanal. Rev.*, 1950, **37**, 101–127.

Tomkins, S. S. *The Thematic Apperception Test.* New York: Grune & Stratton, 1947.

van Lennep, D. J. *Psychologie van projectieverschijnselen.* Utrecht: Ned. Slichting voor Psychotechnick, 1948.

van Lennep, D. J. The four-picture test. In H. H. Anderson & Gladys L. Anderson (Eds.), *An introduction to projective techniques and other devices for understanding the dynamics of human behavior.* New York: Prentice-Hall, 1951.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

MMPI DIFFERENCES BETWEEN PARENTS OF DISTURBED AND NONDISTURBED CHILDREN[1]

SHEPHARD LIVERANT
Ohio State University

It is a widely held clinical notion that the parents of behaviorally disturbed children are themselves more maladjusted than the parents of nondisturbed children. This notion is generally supported by clinical evidence (i.e., social work and psychiatric interviews) and follows from the relatively large amount of research (e.g., Baldwin, Kalhorn, & Breese, 1945; Brown, 1942) demonstrating the effects of parental attitudes and personality upon child behavior. However, the author knows of no studies which directly test this hypothesis by comparing parents of clinically disturbed children with parents of "normal" children in terms of some adjustive criteria.

It is conceivable that intensive study of the parents of nonproblem children may reveal a degree of pathology in their backgrounds similar to that of parents of disturbed children. For example, Goodstein and Dahlstrom (1956) found no significant Minnesota Multiphasic Personality Inventory (MMPI) differences between parents of stuttering and nonstuttering children. The present research is in part an application of their basic design to a different experimental population. It also investigates possible parental syndromes within the disturbed group which may differentially contribute to certain patterns of child pathology, such as acting-out vs. nonacting-out problems.

Subjects and Method of Data Collection

The booklet form of the MMPI was administered as part of the standard diagnostic procedure to all parents bringing a child for evaluation to the Department of Psychiatry at the University of North Carolina Medical School from April, 1956 to June, 1957. In order to be included in the final analysis, both parents had to have taken the test and their child judged by the clinic staff as neither mentally defective nor organically impaired. Forty-nine sets of parents (all Caucasian) met the above criteria.

The comparison group data were drawn from the population of control parents used in the Goodstein and Dahlstrom study (1956).[2] These were parents of essentially physically normal, white children with no known behavioral disturbances. They consisted of volunteers from PTAs, church groups, etc. in and around the Iowa area. In all cases (as with the experimental group) both parents were living together at the time of the study.

Each experimental parent was matched to a control parent on the basis of age, education, rural or urban residence, and, in the case of fathers, socioeconomic status. In order to achieve the closest possible matching, the control mothers and fathers were matched separately with the experimental mothers and fathers rather than by families.

The mean age of the experimental fathers was 40.49 years (range, 30 to 58 years), with average education of 11.59 years (range, 5 to 20 years). That of the control fathers was 40.79 years (range, 30 to 58 years), with average education of 12.88 years (range, 8 to 20 years). The mean age of the experimental

---

[1] The author is indebted to the child psychiatry staff at the University of North Carolina for their cooperation in collecting the data for this study.

[2] A special debt of gratitude is owed Grant Dahlstrom who painstakingly provided the matched sample of control parents' MMPIs from the larger group used in his study.

Table 1

Means, *SD*s, and Significance of Difference Between Means of Experimental and Control Fathers

	Experimental Fathers ($N = 49$)		Control Fathers ($N = 49$)		
Scale[a]	Mean	*SD*	Mean	*SD*	t[b]
?	8.6	28.5	3.0	9.0	1.3
L	4.8	2.2	4.0	2.5	1.1
F	3.6	2.5	3.6	3.3	0.1
K	16.1	5.1	16.1	5.4	0.0
Hs	60	5.0	52	3.4	3.9**
D	61	4.7	53	3.6	3.5**
Hy	62	4.8	56	4.7	3.5**
Pd	60	4.8	56	4.1	1.7
Mf	57	4.4	57	4.6	0.0
Pa	54	2.4	52	2.6	1.4
Pt	57	4.7	52	4.5	2.4*
Sc	53	4.4	52	5.2	0.0
Ma	52	4.1	53	3.8	0.1
A	13.9	9.9	11.2	7.3	1.6

[a] The 9 clinical scales (*Hs* through *Ma*) are in *T*-score units including the *K* correction. The others are raw scores.
[b] All t values are based on raw scores.
* Significant between the .05 and .01 levels of confidence.
** Significant at greater than the .01 level of confidence.

mothers was 36.84 years (range, 26 to 55 years), with average education of 12.37 years (range, 2 to 20 years). That of the control mothers was 37.31 years (range, 27 to 54 years), with average education of 13.16 years (range, 8 to 21 years).

Eighty-two per cent of the experimental families had urban residences, whereas 67% of the control group had urban residences. Over 90% in both groups were Protestants, and they were largely lower middle-class families.

The MMPIs were hand-scored for the four validity scales, nine clinical scales, and the Taylor anxiety scale (1953).

Results

Between Group Comparisons

None of the profiles in either group were excluded on the basis of the validity scale scores. With the exception of two experimental fathers who slightly exceeded a raw score of 7 on the *L* scale, the experimental group met the criteria used by Goodstein and Dahlstrom (1956) to exclude parents whose validity scale scores indicated a significant degree of dissimulation.

Since the *T* scores of the experimental fathers were significantly higher on the *Hs*, *Mf*, and *Ma* scales and significantly lower on the *Pa* and Anxiety (*A*) scales than the experimental mothers, the sexes were not combined in the subsequent comparisons with the control groups. The control fathers had significantly higher *T* scores than the control mothers on the *Mf* and *Ma* scales.

For the sake of reader convenience, the nine clinical scale scores (*Hs* to *Ma*) in Tables 1 and 2 are reported in terms of standard scores (*T* scores) rounded off to the nearest whole number, although all computations are based on the more exact raw scores. As shown in Table 1, the experimental fathers had significantly higher mean scores on the *Hs*, *D*, *Hy*, and *Pt* scales than the control fathers. Table 2 shows that the experimental mothers scored significantly higher on the *F*, *D*, *Hy*, *Pd*, *Pa*, *Pt*, and *A* scales, whereas the control mothers had a significantly larger score on the *K* scale.

The various groups were compared as to severity of disturbance by counting the number of individuals in each group who had one or more *T* scores on the nine clinical scales at

Table 2

Means, *SD*s, and Significance of Difference Between Means of Experimental and Control Mothers

	Experimental Mothers ($N = 49$)		Control Mothers ($N = 49$)		
Scale[a]	Mean	*SD*	Mean	*SD*	t[b]
?	5.9	25.2	1.3	4.3	1.3
L	4.5	2.1	3.7	2.0	1.9
F	4.0	3.6	2.5	1.8	2.5*
K	14.5	4.9	16.9	4.3	2.6**
Hs	54	4.9	52	3.6	1.1
D	59	7.0	52	5.9	2.7**
Hy	60	6.2	56	4.4	2.3*
Pd	61	5.2	53	4.9	2.8**
Mf	48	5.0	49	3.9	1.1
Pa	58	3.7	54	2.3	2.4*
Pt	58	7.5	51	4.7	2.5*
Sc	54	8.9	52	4.8	1.0
Ma	49	3.8	48	2.7	0.3
A	18.9	9.7	12.5	7.3	3.4**

[a] The 9 clinical scales (*Hs* through *Ma*) are in *T*-score units including the *K* correction. The others are raw scores.
[b] All t values are based on raw scores.
* Significant between the .05 and .01 levels of confidence.
** Significant at greater than the .01 level of confidence.

70 or above (the usual cutting score for psychiatric disturbance). Twenty-six of the experimental fathers had one or more T scores at or above 70 as compared to 15 of the control fathers, and 22 of the experimental mothers had one or more scores in this range as compared to only 9 of the control mothers. Chi square = 5.06 ($p < .03$) for the fathers, and chi square = 7.98 ($p < .01$) in the case of the mothers.

It is also significant that there were 10 mothers in the experimental group with 3 or more T scores above 70 and only 2 such mothers in the control group (chi square = 6.10, $p < .02$).

Since present-day MMPI interpretation places a great deal of emphasis on response patterns as well as elevation of scores, the various groups were compared in terms of relative frequency of occurrence of certain two-digit high point codes (i.e., the top two scores). Neither the fathers nor mothers in either group had a frequency in any given code pair exceeding five, and in all four groups there were almost as many different high point code pairs as there were people.

A further analysis of the high point code was made by counting the frequency of occurrence of each clinical scale in either first or second place within each group. In only one instance was a significant difference obtained between experimental and control fathers or between experimental and control mothers. Twelve of the experimental fathers had an *Hs* score in either first or second place as compared to only three of the control fathers. This difference is significant at greater than the .02 level of confidence (chi square = 6.16).

Within Group Analysis

In an attempt to gain some further understanding of the nature of the relationship (if any) between parental pathology and child disturbance patterns, the experimental group was divided into four subgroups on the basis of symptomatology exhibited by the child as follows:

1. The schizophrenic group (S) consisted of parents whose children ($N = 9$) were diagnosed as childhood schizophrenics or severely schizoid (i.e., exhibited bizarre thought processes and/or markedly withdrawn and peculiar behavior).
2. Parents of children ($N = 10$) who had been referred to the clinic because of delinquent or antisocial behavior were placed in the acting-out group (AO).
3. The physical complaint group (PC) was composed of parents whose children ($N = 18$) were suffering from various bodily symptoms for which no physiological basis could be determined (ranging from enuresis to headaches).
4. The neurotic group (N) was constituted of parents of children ($N = 12$) who demonstrated behaviors of a nonacting-out variety (such as difficulties in academic learning) but who were not considered as severely disturbed as the children in Group 1.

Since the parents of male children had no MMPI clinical scales significantly different from the parents of female progeny, sex differences among the children were not considered in composing the above groups. There were 34 males and 15 females ranging in age from 4 to 17 years. Four of the girls were in the schizophrenic group, none in the acting-out group, eight in the physical complainers and three in the neurotic group.

The four validity, nine clinical, and Taylor anxiety scores of each set of fathers in one group were compared with the fathers in each of the other three groups (i.e., S fathers were compared with AO fathers and with PC fathers, etc.). t tests of the difference between means were also computed in the same manner for the mothers, making a grand total of 12 sets of comparisons (14 in a set) or 168 t tests.

Only eight of these t tests reached or surpassed the .05 level of confidence, any or all of which may be significant on a chance basis. However, five of the significant differences were found with respect to the *Mf* scale which may lend some validity to them.

The S mothers were higher (more feminine) on the *Mf* scale than the AO mothers ($t = 2.5$, $p < .05$). Fathers of both schizophrenic and neurotic children had significantly higher *Mf* scores than fathers of acting-out and physical complaint children as follows: between S and AO ($t = 3.7$, $p < .05$), between S and PC ($t = 2.9$, $p.$ $.05$), between N and AO ($t = 2.54$, $p = < .05$), and between N and PC ($t = 2.69$, $p < .05$). The AO fathers had a higher mean *Pd* score than the S fathers which just fails to reach significance at the .05 level. The N fathers were significantly lower on the D scale and significantly higher on the *Ma* scale than the PC fathers ($t = 4.22$, $p < .01$ and $t = 2.37$, $p < .05$, respectively).

On the basis of the number of individuals with one or more T scores at or above 70, the S and N mothers combined were less severely disturbed than the AO and PC mothers (chi square $= 4.13$, $p < .05$). Similar differences were not found in the case of the fathers. A pattern analysis revealed no significant differences between any of the groups.

Discussion

Before interpreting these results the problem of regional differences between the experimental and control parents must be considered. Although previous studies (Welsh & Dahlstrom, 1956) have found various MMPI scales to differ significantly as a function of certain demographic variables, region per se does not appear to have this effect. Goodstein (1954) failed to demonstrate regional differences among large groups of male undergraduates at eight different colleges scattered throughout the United States.

In general, then, these results support the major hypothesis that parents of disturbed children are themselves more disturbed than the parents of nondisturbed children. Of course, this study does not enable us to determine whether the elevated MMPI scores of the experimental parents are related to the etiology of their children's problems or are a result of them or both.

While the mean scores of the experimental parents indicate that they tend to be more maladjusted than adults in general, they certainly do not appear to be as abnormal as adults who seek psychiatric help for themselves. As a matter of fact, the majority have scores more comparable to that of a normal group and as revealed by the pattern analysis they do not fit into the usual clinical categories. This finding should be considered in deciding upon a course of therapy with parents who bring their children for psychiatric or psychological aid.

Because of the large within-group variance on any one MMPI scale, the attributing of psychologically incapacitating characteristics to the experimental parents on the basis of mean scores is at best indicative of certain group trends and does not necessarily apply to any individual within the group. Within the context of the above reservation, the elevated scales of the fathers indicate that they tend to be weak, ineffectual males with an assortment of somatic concerns, whereas the experimental mothers are much more likely to act out their angry and disturbed feelings in relation to others as evidenced by their relatively high *Pd* and *Pa* scores. The major exception to this general characterization occurs in the case of the fathers of acting-out children who tend to have significantly higher *Pd* scores than any other group of males. It would also appear that the experimental mothers demonstrate more disturbed verbal and overt behavior as compared to normal mothers than do the experimental fathers in comparison with the control fathers. In all cases, these findings are consonant with clinical impressions.

Essentially, the present study failed to establish any definitive relationships between parental pathology and the kinds of behavior problems exhibited by their children. This may be a function of the inapplicability of the MMPI for such fine discriminations, the lack of within group homogeneity, the small size of the groups, or any combination of these factors.

One somewhat contradictory finding from a common sense point of view is that the mothers of the schizophrenic and neurotic children (who are ordinarily considered the more severely disturbed children) showed less disturbance on the MMPI than the mothers of acting-out and physically complaining children. An analysis of the socioeconomic background of the parents indicates a tendency for the parents of the schizophrenic and neurotic children to have a somewhat higher socioeconomic status than the parents of the other groups. As a consequence they may be somewhat less prone to answer the MMPI items in a deviant direction, and more likely to inhibit antisocial behavior. They may also be more enlightened concerning the functional basis of some physical disorders and thus less likely to reward them. Cultural differences may also explain the higher *Mf* scores of the S and N fathers.

It is obvious that this study raises more questions than it answers. The next step would be the initiation of a series of investigations to determine the parental behavioral patterns which are the basis for specific childhood problems.

Summary

The MMPI was used to compare 49 sets of parents of clinically disturbed children (experimental group) with 49 fathers and 49 mothers of nondisturbed children (control group). The mean differences between the experimental and control groups on the 14 MMPI scales analyzed and a frequency count of the number of individuals in each group who had one or more clinical scales at or above a T score of 70 (the usual psychiatric cutting score) strongly supported the general clinical impression that both fathers and mothers of disturbed children are themselves significantly more maladjusted than the fathers and mothers of nondisturbed children. However, a pattern analysis and a visual comparison with the typical norms for psychiatrically disturbed adults indicated that the experimental parents are more like "normal" parents than they are like adults who seek psychiatric help for themselves.

An attempt to relate MMPI syndromes within the experimental group to various kinds of behavior problems exhibited by their children essentially met with failure. The differences that were found (primarily on the *Mf* scale) between parents of schizophrenic and neurotic children as compared to parents of acting-out and physical-complaining children seemed to be most parsimoniously accounted for on a cultural basis.

Received May 14, 1958.

REFERENCES

Baldwin, A. L., Kalhorn, Joan, & Breese, Fay H. Patterns of parent behavior. *Psychol. Monogr.*, 1945, **58**, No. 3 (Whole No. 268).

Brown, F. An experimental study of parental attitudes and their effects on child adjustment. *Amer. J. Orthopsychiat.*, 1942, **12**, 224–231.

Goodstein, L. D. Regional differences in MMPI responses among male college students. *J. consult. Psychol.*, 1954, **18**, 437–441.

Goodstein, L. D., & Dahlstrom, W. G. MMPI differences between parents of stuttering and nonstuttering children. *J. consult. Psychol.*, 1956, **20**, 365–371.

Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Welsh, G. S., & Dahlstrom, W. G. *Basic MMPI readings in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

RESPONSES TO THE WORDING OF PERSONALITY TEST ITEMS

CHARLES HANLEY
Michigan State University

The discovery of "acquiescence" in personality tests, having led to considerable research in recent years, seems likely to continue its influence in the near future. Studies have tacitly assumed that the response set is a unitary source of test variance, perhaps because Cronbach (1950), in his influential article on the topic, attributed to acquiescence such diverse behaviors as checking misspelled words on a spelling test, checking "like" on interest inventories, checking "agree" on attitudes, and checking "no" on a temperament schedule. Little attention, however, seems to have been given to the possibility that there might be more than one kind of acquiescence.

Acquiescence may be a personality trait. Leavitt, Hax, and Roche (1955), for example, related the set to "authoritarian submission," and it is certain that the future will bring more studies of the personality correlates of acquiescence. On the other hand, Cronbach (1950) suggested that the set will usually not have predictive value, implying that acquiescence is mainly a function of the test situation itself. Consistent with this belief are assumptions that acquiescence occurs most often when items are ambiguous (Cronbach, 1950), or "controversial" (Fricke, 1956). Accepting the hypothesis that the nature of the test items influences the response set, the present study explores the possibility that there are different "acquiescences," and that these include responses to specific types of item wording.

An early instance of acquiescence was given by Lorge (Lorge, 1937; see Cronbach, 1950 for fuller discussion), who found a general tendency to like (or dislike) items on the Strong Interest Inventory; his long "gen-like" scale had high internal consistency. It is possible that systematic differences in the manner subjects defined the word "like" contributed to the reliability of the dimension. A subject who defined the word strictly ("It must mean *really* like") would, other things being equal, have few likes on the test. One who adopted a loose definition ("Oh sure, I like that") would check many likes. Even though both individuals had identical underlying attitudes toward the activities described in the items, the systematic difference in definition of the key word, "like," would produce "gen-like" behavior.

"Like" is but one of many terms that might elicit systematic differences in interpretation: "often," "wish," etc. are common in personality tests. It also is possible that relative strictness in defining one key word is unrelated to the manner in which others are interpreted. If this occurs, acquiescences, rather than acquiescence, characterize personality inventories. The present study produces evidence on this point.

Procedure

Exploratory analysis of responses to the Minnesota Multiphasic Personality Inventory revealed that internally consistent scales could be defined solely from item stems (e.g., "I often," "I have had"). MMPI items, however, separate deviants from normals, and item content tends to describe personality difficulties. The reliability of the provisional scales thus might be due to item content rather than individual differences in interpretation of wording. For this reason, six new scales were devised, each with 20 items. The investigator, lacking a pool of random meaningful items, attempted to vary content and avoid measuring unitary personality traits. The scales, mixed in a true-false inventory, were given to 59 males and

63 female students in undergraduate psychology courses at Michigan State University.[1] A score on a given scale consists of the number of items marked "true."

The *Like* scale consists of 20 items headed by the words "I like" (e.g., I like work that calls for concentration and carefulness, I like picnics, etc.). Four other scales are headed "I wish," "I never," "I have had," and "I often." The sixth scale, *Global,* differs in that its items have no common stem. Item wording on the *Global* scale was designed to parallel the material in the California F Scale and similar attitude measures, in that self reference is absent, and agreement with broad, or ambiguous, or somewhat grandiloquent attitudes is asked for. It was believed that scores would be determined in part by how critical an *S* was in dealing with items (i.e., critical *S*s would disagree and less critical persons agree). Items ranged from the esoteric (The real rewards of true religious experience come not from earthly prosperity or hope of future salvation, but rather from sublime feelings which defy description with our limited vocabularies) to the more prosaic (It is easier to punish people than to judge them).

Table 1

Internal Consistency of the Scales

Scale	Kuder-Richardson Formula 20 Reliability	
	59 Males	63 Females
I often	.600	.601
I wish	.537	.638
Global	.422	.522
I like	.428	.289
I never	.406	.337
I have had	.393	.096

Results

Data are presented separately for the sexes. The internal consistencies of the six scales, estimated by Kuder-Richardson Formula 20, are shown in Table 1. Moderate reliability characterizes the measures, the exception being *Have Had* in women. The scales are short and have not been improved by item analysis, but the evidence for internal consistency is clear. The reliability of the *Like* scale is

Table 2

Empirical and Theoretical (in Parentheses) Correlations* Among Scales
(Males above, females below diagonal)

Scale		Scale					
		O	*W*	*G*	*L*	*N*	*H*
I often	(*O*)	— —	.21 (.57)	.21 (.50)	.08 (.51)	.06 (.49)	.42 (.49)
I wish	(*W*)	.37 (.62)	— —	.30 (.48)	.03 (.48)	.06 (.47)	.15 (.46)
Global	(*G*)	.25 (.56)	.31 (.58)	— —	.22 (.42)	.21 (.41)	.36 (.41)
I like	(*L*)	.05 (.42)	.35 (.43)	.02 (.39)	— —	.03 (.42)	.24 (.41)
I never	(*N*)	−.07 (.45)	−.22 (.46)	.29 (.42)	−.35 (.31)	— —	−.11 (.40)
I have had	(*H*)	.41 (.24)	.46 (.25)	.24 (.22)	.16 (.17)	−.11 (.18)	— —

* Values of .26 and .25 are significant at the .05 level for males and females, respectively.

---

[1] Copies of the inventory may be obtained from the author.

Table 3

Unrotated First Centroid Factor Loadings and Communalities

Scale	Factor Loading		Factor Communality	
	Men	Women	Men	Women
I often	.534	.489	.285	.239
I wish	.401	.714	.161	.510
Global	.633	.264	.401	.070
I like	.313	.494	.098	.244
I never	.176	−.308	.031	.095
I have had	.565	—	.319	—

quite comparable to Lorge's results (see Cronbach, 1950), provided allowance is made for the difference in number of items. Correlations between scales are given in Table 2.

While systematic responses occurred, results may have been due to an influence common to all six measures. The internal consistency of the *Often* scale, for example, could have arisen because of general acquiescence (or some combination of influences). There are at least two ways to examine this possibility. One of these, factor analysis, is more orthodox but leads to difficulties. First factor loadings of a centroid factor analysis (using each scale's largest correlation as its diagonal entry) are shown in Table 3, together with the communality of the scale attributable to the first factor. The *Have Had* scale in women was not included because of its low internal consistency.

When factor communalities are compared with the corresponding entries in Table 1, it is seen that reliabilities usually are much larger.[2] When reliabilities and communalities are approximately equal, as in the case of the *Global* scale in men, it might seem reasonable to hypothesize a general acquiescence factor. The contrary is true in women, however, where the *Global* scale's first factor communality is much smaller than its reliability. If the reliabilities and communalities are averaged for the two sexes, it will be seen that the first factor accounts for less than half of the reliable variance of the scales.

There is reason to doubt whether even this contribution to scale variance can be attributed to general acquiescence. The first centroid factor can be the average of two or more independent common factors. Further extraction of factors and rotation of axes would be necessary to settle the question. Subjecting the present data to multiple factor analysis, on the other hand, encounters the objection that there are very few tests. Furthermore, if rotation to any criterion were attempted, the general factor loadings would decrease—less evidence of general acquiescence must result. Factor analysis is unsatisfactory for these reasons, but a second simple approach is possible.

The hypothesis being considered may be stated: "The internal consistency of each scale is due to general acquiescence." Since it is assumed that the method of item selection has randomized content factors among the scales, an important deduction follows from this proposition. The scales will correlate with one another, and the magnitude of their intercorrelations will be a function of their reliabilities. When reliabilities can be estimated, as in the present study, it is possible to construct a matrix of correlations expected on the hypothesis that general acquiescence is responsible for internal consistency. Each hypothetical coefficient is given by the formula for the correction for attenuation entered with a "true" value of 1.00 and the relevant reliabilities, and solving for the "empirical" coefficient. Each hypothetical correlation thus is given by the product of the square roots of the reliabilities involved.

These hypothetical coefficients are shown in Table 2 in parentheses below the corresponding actual values. In most cases, empirical correlations are markedly smaller than theoretical coefficients. In view of its internal consistency, the *Wish* scale in men, for example, should have correlations of .57, .48, .48, .47, and .46 with the other scales, if the acquies-

[2] The internal consistency coefficient sets a lower bound for a scale's potential correlation with other measures. To the extent that single items in a scale reliably reflect sources of variance absent in the other items, the measure will have a greater communality with other tests of similar item heterogeneity. In the present study, the use of K-R 20 reliability as a "roof," rather than a "floor," on expected correlations means that the analysis is conservative (i.e., it is biased against the hypothesis of variance specific to item wording).

cence hypothesis is correct. The actual correlations of .21, .30, .03, .06, and .15, respectively, are much lower. It follows that the internal consistency of this scale is too large to have been produced entirely by acquiescence shared with other measures. The alternate hypothesis, the existence of a response set specific to items headed "I wish," is thereby supported. Similar reasoning applies to the other scales, with the exception of the *Have Had* measure in women. This scale has virtually zero internal consistency. The fact that it correlates with the other measures must be attributed to the operation of small independent factors shared with other scales. While correlations between tests have been treated as occurring from general acquiescence, this last finding suggests the presence of influences not contributing to internal consistency but operating to produce correlation between measures that on first sight might seem the result of general acquiescence.

ACE scores were available for 36 males and 49 females. One correlation with the scales proved statistically significant (.05 level), −.34 between ACE and the *Global* scale in males (the corresponding value for females was −.16). This result lends some support to the assumption that responses to Global items are partly a function of the criticalness with which *S*s respond, if criticalness is assumed to involve some intellectual ability.

When the six measures were pooled in a "general acquiescence" measure, the K-R 20 reliabilities were .73 for males and .70 for females. On over-all reliability alone, therefore, one could hypothesize a general acquiescence factor, whereas finer-grain analysis would show independent sources of reliable variance.

Discussion

If the findings of this exploratory study are either accepted or substantiated by further work with other stems and content, a number of important consequences follow. Studies seeking to correct personality scales for acquiescence should attend carefully to their acquiescence measures. The same caution is necessary when investigators set out to find the characteristics of the "acquiescent personality." It is entirely possible that many acquiescences exist. Some of these could reflect group factors. There might be, for example, relatively high correlations between scales defined by "often," "usually," "at times," and "most of the time." If such a "temporal" set exists, it could have interesting personality correlates. To search for a single set of correlates of acquiescence, however, would be to disregard the complexity of test-taking behavior, and to lead to an oversimplified view of human nature.

When a scale contains a variety of item stems, each eliciting individual differences in interpretation, the result is to lower the internal consistency of the measure. Thus, when different item wordings are mixed in a single scale, a certain amount of "noise," as opposed to "information," is to be expected. To make items homogeneous in style, on the other hand, is to inflate the internal consistency of the scale. Such cases occur and are not limited to true-false tests. Edwards (1954), for example, has worded the bulk of the items on his forced-choice Personality Preference Schedule in terms of "liking" something. The exceptions are all items on the Abasement and the majority on the Aggression scales. When both items in a forced-choice pair involve "liking," the set is nullified, but when an abasement item is always paired with a "liking" item, the set could operate. It seems reasonable to expect that new forced-choice personality tests will appear in the future; consequently, it is to be hoped that strict attention is paid to item wording before interpretation of item content is attempted.

Summary

Six 20-item true-false scales were constructed to test the hypothesis that responses on personality tests are in part a function of the subject's interpretation of the wording of items. Five of the scales were homogeneous for item stems: "I often," "I wish," "I have had," "I never," and "I like." The sixth scale contained varied content worded in the manner of California F test items.

The measures showed moderate reliability when given to groups of male and female university students. Intercorrelations among scales generally fell below the values to be expected if the reliability of the individual

measures were due solely to general acquiescence. There appear to be different types of acquiescent behavior, determined in part by the interpretation subjects make of key words in items.

It was concluded that studies of "acquiescence" in tests, and investigations of the "acquiescent personality," require considerable attention to item wording before results can be accepted as definitive.

Received May 20, 1958.

REFERENCES

CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt,* 1950, **10**, 3–31.

EDWARDS, A. L. *Personality Preference Schedule, test manual.* New York: Psychological Corp., 1954.

FRICKE, B. G. Response set as a suppressor variable in the OAIS and MMPI. *J. consult. Psychol.,* 1956, **20**, 161–169.

LEAVITT, H. J., HAX, H., & ROCHE, J. H. "Authoritarianism" and agreement with things authoritative. *J. Psychol.,* 1955, **40**, 215–221.

LORGE, I. "Gen-like:" Halo or reality? *Psychol. Bull.,* 1937, **34**, 545–546.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

THE STABILITY OF TAT FANTASY AND STIMULUS AMBIGUITY[1]

JEROME KAGAN
Fels Research Institute, Yellow Springs, Ohio

Although psychologists use fantasy responses for both research and diagnostic purposes, information on the long-term stability and validity of fantasy variables is still needed. This report deals with the long-term stability of certain TAT content categories in relation to the ambiguity or "content pull" of the fantasy stimulus. Ambiguity is defined in terms of the number of alternative interpretations given to a stimulus and as ambiguity decreases one expects several popular or modal interpretations to be elicited by the stimulus. When a fantasy stimulus is highly ambiguous, individual differences in thematic content may be due, in part, to differences in (*a*) the interpretation of ambiguous cues on the stimulus and (*b*) the subject's cognitive approach to the task (Schwartz & Kates, 1957; Weisskopf, 1950; Weisskopf-Joelson & Lynn, 1953). For example, the thematic content produced to Cards 11 or 19 of the TAT series, which are highly ambiguous, seems dependent on both the subject's hypothesis as to the nature of the "object" on the picture and his cognitive style in interpreting an ambiguous stimulus. Some people become descriptive of concrete details of the stimulus when asked to interpret an ambiguous picture, while others, perhaps less anxious, project motives and feelings into their themes (Weisskopf-Joelson & Lynn, 1953).

However, when the stimulus is structured to suggest a specific motive or conflict, one may assume that one of the subject's hypotheses for interpretation included the motive illustrated. Thus, absence of a certain motive in a theme told to a stimulus suggests that content is apt to be a more valid measure of conflict than absence of this content to an ambiguous stimulus. For example, absence of an aggressive theme to a picture with no cues for aggression could indicate either (*a*) low aggressive motivation, (*b*) strong aggression anxiety which led to inhibition of an aggressive theme, or (*c*) a descriptive and nonthematic approach to ambiguous stimuli. However, when a subject does not report an aggressive theme to a picture possessing aggressive cues, it seems most plausible to account for this behavior in terms of anxiety over aggressive motives. Thus, if conflict remains stable over time, occurrence or nonoccurrence of thematic content to stimuli which suggest conflictful motives is apt to be more reliable than the interpretations of ambiguous stimuli. Recent research suggests that fantasy interpretations of stimuli which suggest specific motives are also better predictors of related classes of overt behavior than fantasy responses to stimuli which are ambiguous for these motive areas (Crandall, 1955; Epstein & Smith, 1956; Kagan, 1956; Kagan & Moss, 1959; E. J. Murray, 1958; Scott, 1956; Weisskopf, 1950; Weisskopf-Joelson, Asher, Albrecht, & Hoffman, 1957).

The present research investigated the relationship between the stability of TAT content and stimulus ambiguity for a group of middle-class children who are part of the Fels longitudinal research population.

Method

From the files of the Institute, 86 cases (*S*s) were selected (44 boys and 42 girls) for whom three sets of seven card TAT proto-

---

[1] This research was supported by a research grant (M-1260) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The author wishes to thank Howard A. Moss for his comments on the manuscript.

cols were available. The median ages of administration were 8–9, 11–6, and 14–6. The seven cards administered from the H. A. Murray (1943) series were 1, 5, 14, 17 BM, 3 BM, 6 BM, and 3 GF.

TAT Content

The TAT protocols were scored for a variety of themes, but many of the themes were of infrequent occurrence and this report deals with the following content categories which have been classified as needs or press.

a. Dependency: (*n-Dep*). The central figure actively seeks aid, support, guidance, or encouragement from other figures.

b. Nurturance: (*p-Nurt*). The central figure is given love, advice, a gift, or encouragement from another figure and *the nurturance was unsolicited by the central figure.*

c. Achievement: (*n-Ach*). The behavior or thoughts of the central figure are concerned with competition with a standard of excellence. This content was scored according to the definition of McClelland (Atkinson, Clark, & Lowell, 1953) and is comparable to their score of *Ach Th.*

d. Physical aggression by the central figure: (*n-Ph Agg*). The central figure executes or contemplates physical aggression, e.g., to fight, kill, injure.

e. Nonphysical aggression by the central figure: (*n-Ind Agg*). The central figure contemplates or executes nonphysical aggression such as teasing, stealing, being angry, etc. This category includes nonphysical aggression at all figures except parents.

f. Nonphysical aggression to the parents by the central figure: (*n-Ind Agg Par*). The central figure contemplates or executes nonphysical aggression toward the mother, father, or both parents.

g. Physical aggression directed at the central figures: (*p-Ph Agg*). The central figure is killed, injured, or physically attacked by another person.

h. Nonphysical aggression directed at the central figure: (*p-Ind Agg*). The central figure is teased, rejected, or robbed by figures other than the parents.

Percentages of agreement between two independent raters scoring the protocols was 88%. The central figures were designated before the scoring in the following manner. For Cards 1, 17 BM, 3 BM, 3 GF, and 14 the single figure on the card was the central figure. For Card 5 the woman was the central figure unless the woman was designated as a mother and her child was a major figure in the plot. In these cases the child was the central figure. For Card 6 BM the young man was designated as the central figure.

Results

Age and Sex Differences in TAT Content

Table 1 shows the percentage of boys and girls reporting one or more of the eight content categories for each of the three protocols.

There were only two marked age changes in frequency of the various contents. In comparing frequency of occurrence for Protocol 3 versus Protocol 1, achievement themes showed a significant increase with age for both sexes ($p < .01$), and the girls showed an increase in physical aggression by the central figure ($p < .01$). All p values are for two tails. There were two significant sex differences in content. For both Protocols 1 and 2 the occurrence of both dependent and nurturant themes were more frequent for girls than for boys ($p < .02$ for both contents). In contrast with earlier findings on the content of

Table 1

Percentage Reporting TAT Themes to Each of Three Protocols

	Protocol 1		Protocol 2		Protocol 3	
Content	Boys	Girls	Boys	Girls	Boys	Girls
n-Dep	15.9	42.9*	13.6	38.1*	20.5	35.7
p-Nurt	15.9	42.9*	15.9	40.5*	20.5	26.2
n-Ach	40.9	35.7	47.7	45.2	59.1	59.5
n-Ph Agg	27.3	14.3	31.8	40.5	27.3	40.5
n-Ind Agg	45.5	28.6	45.5	40.5	27.3	40.5
n-Ind Agg Par	29.5	31.0	11.4	38.1*	22.7	40.5
p-Ph Agg	27.3	33.3	22.7	16.7	22.7	19.0
p-Ind Agg	22.7	14.3	31.8	35.7	15.9	31.0

* Sex differences significant at $p < .02$ for two tails.

doll play fantasy in preschool children, there were no marked sex differences in incidence of aggressive fantasy (Bach, 1945; Sears, 1951; Yarrow, 1948).

Stability of TAT Content

Since the average occurrence of each of the content categories for any one protocol was one, phi coefficients were computed to assess the stability of the eight content categories over the three administrations. There were no marked sex differences with respect to the degree of stability of any of the content categories and the sexes were pooled in computing the final phi coefficients. Table 2 shows the values of phi and associated two-tailed significance values.

Table 2

Reliability of Occurrence of TAT Themes Over Three Administrations (Phi Coefficients)

Content	P1–P2	P2–P3	P1–P3	Average Phi
n-Dep	+.18	+.01	+.01	.07
p-Nurt	+.01	+.03	+.10	.05
n-Ach	+.32*	+.16	+.22	.23
n-Ph Agg	+.34*	+.34*	+.15	.28
n-Ind Agg	+.18	+.16	+.17	.17
n-Ind Agg Par	+.01	+.11	+.20	.11
p-Ph Agg	+.09	+.14	+.19	.14
p-Ind Agg	+.01	+.27*	+.02	.10

* $p < .01$ for two tails.

It is apparent that, with two exceptions, the content dimensions were not very stable over time.[2] The two categories which showed a low but significant degree of stability were physical aggression by the hero and need achievement. Degree of stability was not simply related to frequency of occurrence, for n-Ph Agg was not one of the more frequently occurring themes. Although p-Ind Agg showed a significant phi coefficient between Protocols 2 and 3 the other two phi coefficients were nonsignificant and this category was not as stable as n-Ach or n-Ph Agg. An analysis of the distribution of all eight contents to the seven TAT stimuli revealed that n-Ach and n-Ph Agg themes each occurred with a high degree of frequency to a specific picture which seemed suggestive of these contents. Physical aggression by the central figure tended to occur to Card 3 BM, which illustrates a slouched figure and an object often seen as a gun. Achievement themes occurred to Cards 1 (boy with violin) and 17 BM (man on rope), both of which suggest competition and/or a desire to accomplish a task or improve at a skill. Occurrence of the remaining six contents, which were less stable, tended to be more evenly distributed over many of the TAT pictures. That is, there was no one stimulus which was highly suggestive of any

Table 3

Percentage of Entire Group Reporting TAT Themes on One or More of the Three Administrations

	Card						
Content	1	17 BM	3 BM	3 GF	14	6 BM	5
n-Dep	12.7	12.7	16.3	16.3	11.6	16.3	12.7
p-Nurt	11.6	26.7	29.0	20.9	5.8	20.9	9.3
n-Ach	59.3	44.2	3.4	0.0	22.0	9.3	5.8
n-Ph Agg	0.0	6.8	44.2	4.7	5.8	8.2	5.8
n-Ind Agg	3.4	12.7	23.2	20.9	20.9	32.6	16.3
n-Ind Agg Par	13.6	1.2	12.7	3.4	18.6	25.4	10.5
p-Ph Agg	0.0	15.1	23.2	24.4	10.5	2.4	4.7
p-Ind Agg	2.4	8.2	15.1	18.6	3.4	8.2	16.3

---

[2] In order to determine if the stability of aggressive themes would change if the categories were pooled, phi coefficients were computed for the combined category of all aggressive themes by the central figure (Categories d, e, f) and for all aggressive themes directed at the central figure (Categories g, h). The stability of these two combined categories was not statistically significant and did not differ from the stability of any of the individual aggression categories except for physical aggression by the central figure.

Table 4

Degree to Which the Occurrence of the TAT Themes Across TAT Stimuli Deviated From a Random Distribution (Values of Cochran's Q)

	P-1	P-2	P-3	Average Q
n-Dep	10.5	5.2	20.2	12.0
p-Nurt	22.3	11.6	6.2	13.3
n-Ach	110.7*	81.6*	92.4*	94.9
n-Ph Agg	35.2*	81.5*	93.1*	69.9
n-Ind Agg	23.4*	11.7	10.0	15.0
n-Ind Agg Par	25.7*	6.0	19.2	17.0
p-Ph Agg	36.3*	27.4*	29.3*	31.0
p-Ind Agg	9.6	10.8	30.0*	16.8

* A value of Q equal to or greater than 22.5 is required for significance at the .001 level of confidence for two tails.

of these six contents. Table 3 shows the distribution of each of the eight content categories to the seven TAT stimuli. The numbers refer to the percentage of the entire group reporting the specific content on at least one of the three administrations.

Table 3 reveals that the only stimuli to which at least 40% of the group told the same theme were Cards 3 BM (n-Ph Agg), 1 (n-Ach), and 17 BM (n-Ach). In order to test the statistical significance of the tendency for physical aggression and achievement themes to be nonrandomly distributed over the seven TAT cards, Cochran's Q test was applied to the data for each of the three protocols. This statistic tests the degree to which the pattern of occurrence of a fantasy theme to the seven TAT stimuli deviated from a random distribution, and the distribution of Q approximates that of chi square. The reader is referred to Cochran (1950) and Siegel (1956) for a summary of this technique. Table 4 shows the values of Q for each content category to each of the three protocols. The larger the value of Q the less random was the distribution of the theme to the seven TAT cards.

The largest values of Q occurred for achievement and physical aggression themes and Q was significant at the .001 level for these two contents for all three protocols. Furthermore, the rank order correlation between the average value of phi (measure of stability) and the average value of Q over the three administrations was +.81 ($p < .05$). This indicates that even for the contents which did not show significant values of phi, the degree of stability of a theme was facilitated by the presence of a stimulus which had some tendency to elicit the content in question. One implication of these findings is that when a fantasy motive or press occurs to a stimulus which does not suggest that content, the long-term stability of that fantasy variable tends to be very low.

Discussion

These data suggest that the long-term stability of a fantasy motive or press is dependent on the ambiguity of the fantasy stimulus for that content. If one assumes that these content variables have some relation to comparable motive and response tendencies in other behavioral contexts, then one implication of this finding is that a theme told to an ambiguous stimulus may be a less sensitive indicator of a motivational predisposition than a theme told to a nonambiguous stimulus. Several investigations support this hypothesis. Kagan (1956) found that boys who were rated as extremely nonaggressive were less apt to tell aggressive stories to pictures which suggested boys fighting than boys who were rated as extremely aggressive. The incidence of aggressive stories to ambiguous pictures did not differentiate these two groups. It was assumed that the low aggression boys were anxious over expression of aggression both in fantasy and overt behavior and inhibited this class of response in both situations. E. J. Murray (1958) found that occurrence of sleep themes to pictures which were ambiguous for this theme did not differentiate sleep deprived from nondeprived *S*s. The stimulus which suggested sleep activity did differentiate the two groups. The sleep deprived *S*s were less likely to tell sleep themes to the nonambiguous stimulus, and it was assumed that sleep deprivation led to anxiety over the thought of sleep and produced an inhibition of sleep related fantasy. Epstein and Smith (1956) found similar results using hunger motivation as the experimental variable. Incidence of hunger themes to pictures ambiguous for this content did not differentiate hungry from nonhungry *S*s. However, to the stimulus which suggested a hunger theme, the hungry men were less apt than the nonhungry *S*s to report this content. The authors suggested, as did Kagan (1956) and E. J. Murray (1958), that when a fantasy stimulus suggests a content which is associated with anxiety, the *S* tends to inhibit a theme which includes the anxious material. This reasoning is based on the interaction of the gradients of approach and avoidance postulated by Miller (1948). Aggression and achievement are both motives which are potential sources of anxiety (Kagan, 1956; Kagan & Moss, 1959; McClelland et al., 1953; Weisskopf-Joelson et al., 1957), and thus the present results are consistent with previous theorizing and research.

One implication of these studies is that a fantasy response to a stimulus which suggests a conflictful motive may be a more sensitive indicator of anxiety or conflict than the response to a stimulus which is ambiguous for that content. Lazarus (1953) has argued for the inclusion of nonambiguous stimuli in projective instruments as a means of assessing conflict and type of defense to conflictful motives. However, there has been a long standing tendency to use projective tests which contained ambiguous stimuli because they were assumed to elicit the primary preoccupations of the *S*. With the recent emphasis on the importance of the defensive organization of the individual and the ways in which he can repress, deny, and distort anxiety arousing motives it seems profitable to use interpretations of nonambiguous stimuli as indices of conflict and associated defensive responses. The present results indicate that interpretations of nonambiguous stimuli are more stable over time and suggest that psychologists evaluating fantasy should take into account the specific stimulus to which the fantasy appears.

Summary

The long-term stability of eight fantasy categories on the TAT was assessed on a group of 86 children with TAT cards administered at ages 8–9, 11–6, 14–6. The eight contents were n-Dependence, p-Nurturance, n-Achievement, n-Physical Aggression, n-Indirect Aggression, n-Indirect Aggression to Parents, p-Physical Aggression, and p-Indirect Aggression. Of the eight contents, only physical aggression by the hero and achievement themes showed better than chance stability over the three administrations. Analysis of the pictures to which these occurred revealed that these two contents occurred to pictures which were suggestive of these contents. The remaining content categories tended to occur with equal frequency to several of the cards and were not frequently elicited by any one stimulus. Thus, the stability of the content was related to the presence of a stimulus which had a tendency to elicit the theme in question. The sensitivity of nonambiguous stimuli in assessing degree of conflict and defensive reactions was discussed.

Received May 21, 1958.

REFERENCES

Bach, G. R. Young children's play fantasies. *Psychol. Monogr.*, 1945, **59**, No. 2 (Whole No. 272).

Cochran, W. G. The comparison of percentages in matched samples. *Biometrika,* 1950, **37**, 256–266.

Crandall, V. J. An investigation of the specificity of reinforcement of induced frustration. *J. soc. Psychol.*, 1955, **41**, 311–318.

Epstein, S., & Smith, R. Thematic apperception as a measure of the hunger drive. *J. proj. Tech.*, 1956, **20**, 372–384.

Kagan, J. The measurement of overt aggression from fantasy. *J. abnorm. soc. Psychol.*, 1956, **52**, 390–393.

Kagan, J., & Moss, H. A. Stability and validity of achievement fantasy. *J. abnorm. soc. Psychol.*, 1959, **58**, 357–364.

Lazarus, R. S. Ambiguity and non-ambiguity in projective testing. *J. abnorm. soc. Psychol.*, 1953, **48**, 443–445.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

Murray, E. J. Thematic apperception during sleep deprivation. Paper read at meeting of Eastern Psychological Association, Philadelphia, 1958.

Murray, H. A. *Thematic Apperception Test manual.* Cambridge: Harvard Univer. Press, 1943.

Mussen, P. H., & Scodel, A. The effects of sexual stimulation under varying conditions on TAT sexual responsiveness. *J. consult. Psychol.*, 1955, **19**, 90.

Schwartz, F., & Kates, S. L. Rorschach performance, anxiety level and stress. *J. proj. Tech.*, 1957, **21**, 154–160.

Scott, W. A. The avoidance of threatening material in imaginative behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 338–346.

Sears, Pauline S. Doll play aggression in normal young children. *Psychol. Monogr.*, 1951, **65**, No. 6.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Weisskopf, Edith A. An experimental study of the effect of brightness and ambiguity on projection in the Thematic Apperception Test. *J. Psychol.*, 1950, **29**, 407–416.

Weisskopf-Joelson, Edith A., & Lynn, D. B. The effect of variations in ambiguity on projection in the Children's Apperception Test. *J. consult. Psychol.*, 1953, **17**, 67–70.

Weisskopf-Joelson, Edith A., Asher, E. J., Albrecht, K. J., & Hoffman, M. L. An experimental investigation of "label-avoidance" as a manifestation of repression. *J. proj. Tech.*, 1957, **21**, 88–93.

Yarrow, L. J. The effects of antecedent frustration on projective doll play. *Psychol. Monogr.*, 1948, **62**, No. 6.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

A STATISTICAL NOTE ON RACIAL DIFFERENCES IN THE PROGRESSIVE MATRICES

ARTHUR R. JENSEN[1]
Institute of Psychiatry, University of London

An article appearing in this Journal (Sperrazzo & Wilkins, 1958) presented an analysis of variance of scores on Raven's Progressive Matrices. The main variables in the analysis were age, sex, race, and socioeconomic (S-E) status (as judged from father's occupation). Variance estimates attributable to age, race, and S-E status were found to be significant beyond the .01 level. However, the fact that there was a significant first-order interaction between race and S-E status and a significant third-order interaction between race, sex, age, and S-E status led the authors to conclude the following:

> It is apparent from the significant race by socioeconomic status interaction and the third-order interaction that a restriction on the interpretation of the race difference found is necessary. The measured differences in scores between races are related to the age, sex, and socioeconomic status of the *S*s. The results cannot be interpreted, therefore, as showing differences in intelligence between the races tested here. The differences found seem to depend upon variations of the nonrace factors (Sperrazzo & Wilkins, 1958, p. 37).

This is an incorrect interpretation of the analysis and, therefore, from a statistical standpoint, an unwarranted conclusion. The most stringent test of the significance of the main effects is obtained in this case, not by using the residual (or within group) variance as the error term, but by including *all* the interactions (in addition to the within group variance) in the error term. This has been done in Table 1.

It may be concluded that in spite of the interactions between the variables (race, S-E status, age, sex), the race difference is highly significant statistically. The approximate percentage of the total variance in Matrices scores in this study attributable to age is 23%; to race, 10%; to S-E status, 5%; and to sex, 0%.

It should also be pointed out that while the Matrices are regarded as a test of general intelligence, with a G saturation of about .80 (Raven, 1947), the term "intelligence" in the above quotation would seem to call for inverted commas in view of the unusually low correlations found in this study between the Matrices and three other tests of intelligence.

Table 1

Analysis of Variance of Progressive Matrices

Source of Variation	Sums of Squares	*df*	Mean Variance	*F*
Total	20,759.96	479		
Age	4,772.60	4	1,193.15	44.13*
Sex	80.03	1	80.03	2.96
Race	2,116.80	1	2,116.80	78.28*
Socioeconomic	1,056.52	2	528.26	19.54*
Residual (error)	12,734.01	471	27.04	

* $p < .001$.

Of course, the actual magnitude of the racial difference may be very small, even though the difference is highly significant statistically. An analysis of variance performed on the three socioeconomic groups separately, in addition to presenting the mean scores for each race within each S-E group, would give a more complete and meaningful picture.

Received June 3, 1958.

REFERENCES

Raven, J. C. *Progressive Matrices (1947), Sets A, Ab, B.* London: H. K. Lewis, 1947.

Sperrazzo, G., & Wilkins, W. L. Further normative data on the Progressive Matrices. *J. consult. Psychol.*, 1958, 22, 35–37.

---

[1] USPHS Research Fellow of the National Institute of Mental Health.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

RACIAL DIFFERENCES ON PROGRESSIVE MATRICES

GERALD SPERRAZZO
University of Ottawa

AND WALTER L. WILKINS
St. Louis University

The additional analyses suggested by Jensen have been undertaken, and the results are shown in Table 1. Three analyses of variance are presented by socioeconomic status of 480 children, this status being based only upon occupation of the father. The three levels of such status are represented as follows: S-E (A): professional or semiprofessional; S-E (B): skilled, semiskilled, or clerical; S-E (C): service, unskilled, or labor. The findings in Table 1 show a significant difference between the races (Negro and white) in two of the three socioeconomic status interactions reported earlier. Mean score for each race at each of these three levels is shown in Table 2.

The authors recognize that the index of socioeconomic status used for both white and Negro groups does not equate the samples from the two races on socioeconomic level as this concept is defined in conventional treatments of social classes. Rather, the index was intended to provide an ordering within each race that would parallel the order obtained from more precise indices. The assertion that American Negro class structure is not directly comparable to that of American white has convincing support from social science literature. Therefore, it is not surprising that "race differences" in "intelligence" test scores are found with a test that is sensitive to variations in socioeconomic background. What is more surprising is that a relatively nonverbal

Table 1

Analysis of Variance of Scores From 480 Children on Colored Progressive Matrices by Socioeconomic Status

Sources of Variation	*df*	S-E (A) Mean Square	S-E (A) *F*	S-E (B) Mean Square	S-E (B) *F*	S-E (C) Mean Square	S-E (C) *F*
Total	159						
Age	4	437.58	17.94*	407.44	16.71*	415.20	14.68*
Race	1	980.10	40.20*	1,452.00	59.58*	105.62	3.73
Sex	1	5.60	—	.90	—	198.02	7.00*
Error	153	24.38		24.37		28.28	

* .01 level of confidence.

Table 2

Mean Scores of Negro and White Children on Colored Progressive Matrices by Socioeconomic Status

Race	S-E (A)[a]	S-E (B)[b]	S-E (C)[c]
White	26.41	24.78	21.13
Negro	21.46	18.76	19.51

[a] Professional, semiprofessional.
[b] Skilled, semiskilled, clerical.
[c] Service, unskilled, laborer.

performance test is demonstrably sensitive to differences in socioeconomic level.

The usefulness of the S-E index even to rank-order Negro class structure is thrown into doubt by an examination of mean differences by S-E index, race, and sex. For white males and females, there is a difference of over five points in mean test score between the top and bottom S-E categories ($P < .01$). Although there is a mean difference of 3.68 between top and bottom S-E category for Negro females ($P < .01$), there are no demonstrated differences among the categories for Negro males ($P > .10$). One may conjecture after the fact that, for example, service occupations (S-E Level C) are functionally equivalent to or higher among Negroes than semiskilled or clerical occupations (S-E Level B). The differences between Negro males and females in the ordering of test scores for the three S-E groups might reflect a sex difference in accuracy of reporting father's occupation or, more precisely, a sex difference in willingness to report a socially less acceptable family arrangement.

The question raised by Jensen leads to speculation on several fronts, and we wish to thank him for stimulating this re-examination.

Received February 2, 1959.

REFERENCES

Jensen, A. R. A statistical note on racial differences in the progressive matrices. *J. consult. Psychol.*, 1959, **23**, 272.

Sperrazzo, G., & Wilkins, W. L. Further normative data on the progressive matrices. *J. consult. Psychol.*, 1958, **22**, 35–37.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

DEFENSIVENESS, INSIGHT, AND THE *K* SCALE

EWART E. SMITH[1]

Fels Group Dynamics Center, University of Delaware

Defensiveness is a popular, but difficult to measure, concept. Page and Markowitz (1955) have defined it as follows: "The defensive individual is described as one who fails to ascribe to himself characteristics of a generally valid but socially unacceptable nature." As Hogan (1952) has pointed out in his theory of threat and defense, there are two general theories of pathology. The first theory holds that pathology is due to frustration, or blocked needs, and the second that pathology is due to threat. As Hogan defines it, "Threat occurs when the self perceives its experiences as being inconsistent with the structure of the self." It is then necessary for the individual, in order to maintain his perceived self, to distort experience, particularly that experience which is contrary to his self image. Consequently, we would expect defensive individuals to be low in self insight.

The *K* scale (Meehl & Hathaway, 1946), which is generally considered a measure of defensiveness (McKinley, Hathaway, & Meehl, 1946; Page & Markowitz, 1955; Sarason, 1956; Grayson & Olinger, 1957; Handley, 1957) because it detects abnormals who described themselves as normal on the MMPI, was used as a defensiveness measure in the present investigation.

The hypothesis tested was: The more defensive the individual, the less insight he will have. Unexpected results necessitated a second study, designed to retest this hypothesis and, in addition, to test the hypothesis that the *K* scale is a measure of defensiveness.

Method

Subjects and Task

In the first investigation the subjects (*S*s) were 45 male and 45 female summer school students at the University of Delaware, who participated in nine male and nine female five-person groups. The task was to solve, by means of a 25-minute discussion, an interesting but difficult industrial personnel problem (Maier, 1950).

In the second investigation, the *S*s were 11 male and 3 female first-line supervisors (foremen), and 12 male second-line supervisors in a large eastern corporation. The *S*s were participating in a five-day management training course in which the interpersonal factors in management were stressed through the use of role-playing, feedback sessions, etc. Two training groups were used, separated in time, one group for first-level and one for second-level supervisors.

Measures

In the first investigation, the *K* scale was used to measure defensiveness. Insight was measured by having each *S* indicate for three negative roles (aggressing, blocking, withdrawing) who (including himself) took the roles. An insight score was computed for each *S* by determining the discrepancy (ignoring the sign) between the number of roles he credited to himself, and the average number of roles credited to him by the group.

In the second investigation, the *K* scale was again administered, as was the role list, expanded to include four negative roles. In addition, each *S* rated each group member on defensive behavior. Defensiveness was defined on the rating form as follows:

> By defensiveness, we mean the tendency of a person to become upset if anyone finds fault with them, the tendency to deny having made an error or mistake when it is obvious that we have made a mistake, the avoidance of evaluations which might be critical, etc.

---

[1] The author wishes to thank Jacqueline Goodchilds and Jacob Schonfield for their helpful comments and criticisms.

Table 1

Number of *Ss* Accurate or Inaccurate on Negative Roles by *K* Scores

	Accurate	Inaccurate	Totals
Above mean on *K* scale	19	25	44
Below mean on *K* scale	9	37	46
Total	28	62	90

$\chi^2 = 5.85^*$

Results [2]

The First Investigation

The *K* scale scores were broken at the mean, and the insight scores were dichotomized into completely accurate or not completely accurate categories. We see from Table 1 that there is a significant relationship between defensiveness, as measured by the *K* scale, and insight, with the more defensive individuals having the most insight. This result is contrary to the hypothesis.

The Second Investigation

The rank order correlation between the *K* scale and insight on negative role behavior is −.52**, which cross-validates the relationship found between the *K* scale and insight in the first investigation. The rank order correlation between the defensiveness rating and insight on negative role behavior is +.60**, which supports the original hypothesis.

The reliability of the defensiveness rating scores, computed for each *S* by averaging the ratings of the other group members, was computed by correlating the mean ratings of the odd numbered and even numbered raters. The resulting corrected reliability correlation is .91. The corrected reliability coefficient on the average negative role scores used for computing the insight scores was .87. Finally, we find that the correlation between the *K* scale and the defensiveness ratings is −.39*.

Discussion and Conclusions

At this point, our results are both persistent and contradictory. Our two measures of defensiveness are negatively correlated. The *K* scale was carefully constructed empirically, and cross-validated. The defensiveness behavioral ratings were made by combining the judgments of 11 or more *Ss*, each of whom had had 40 hours of observation on which to base their ratings.

It is possible to rationalize the data, however, if it is recalled that the *K* scale items are essentially a list of negative behaviors which the *S* is asked to either accept or deny as applying to himself. Consequently, when the original *K* scale standardization *Ss*, who were abnormal on such criteria as institutionalization, obtained high scores on the *K* scale (i.e., denied abnormal behavior) it was clear that they were being defensive. However, in the two investigations reported here, we were dealing with normal *Ss* and found that normal *Ss* who denied, on the *K* scale, considerable abnormal behavior had good insight and were not seen as behaving defensively. On the other hand, our normal population *Ss*, who admitted a large number of negative traits on the *K* scale, had poor insight and behaved defensively, suggesting some degree of insecurity or neuroticism on their part. In sum, it is argued here that it is defensive for abnormal population *Ss* to obtain high *K* scale scores but a sign of health for normal populations *Ss*. This explanation can be supported by several other investigations. Wheeler, Little, and Lehner (1951) in their factor analysis of the MMPI present evidence for an interpretation of the *K* scale as a measure of ego defenses. Sarason (1956), using the *K* scale as a measure of defensiveness, found that those scoring high in defensiveness were significantly superior on the ACE and tended to be superior in grade-point average. These results were interpreted as due to the adverse effect on performance of the self-criticism of low *K*s. Sweetland and Quay (1953) summarize several bits of information which indicate that normal *K* scores may indicate social security and emotional adjustment. They conclude: "The suggestion was made that *K* scores, when within normal ranges, may possibly be a measure of personality integration." In view of the evidence, as seen in Wheeler, Little, and Lehner's data, that the mean *K* scores of normal population groups are higher than those of abnormal populations, Sweetland and Quay might better have concluded that high *K* scores *on normal populations* are a measure of personality integration.

---

[2] Statistical significance at the .05 level has been indicated by a single asterisk; at the .01 level by a double asterisk.

In conclusion, it is believed that the data supports the hypothesis that defensive individuals are low in insight. Secondly, the data indicates that the *K* scale should not be used as a measure of defensiveness with normal populations. In addition, these results suggest that greater caution should be exercised in the intepretation of clinically derived measures used with normal populations.[3]

Summary

Two investigations were made of the relationship between defensiveness and insight. The hypothesis tested was that defensive individuals would be low in insight. In the first (laboratory investigation), using the *K* scale of the MMPI as a measure of defensiveness, a significant relationship was found but was not in the predicted direction.

In a second (field investigation), the original hypothesis was supported by a high positive relationship between defensive behavior ratings and insight. The *K* scale again correlated in the wrong direction with insight and was negatively related to the defensive behavior rating.

It was concluded that: 1. Individuals who behave defensively in groups are low in insight. 2. The *K* scale is not an adequate measure of defensiveness on normal populations. 3. Greater caution should be exercised in the use of clinically derived measures on normal populations.

Received November 17, 1958.
Early Publication.

REFERENCES

Grayson, H. M., & Olinger, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI. *J. consult. Psychol.*, 1957, **21**, 73–77.

Handley, C. H. Deriving a measure of test-taking defensiveness. *J. consult. Psychol.*, 1957, **21**, 391–397.

Hogan, R. A. A theory of threat and defense. *J. consult. Psychol.*, 1952, **16**, 417–424.

Maier, N. R. F. The quality of group decisions as influenced by the discussion leader. *Hum. Relat.*, 1950, **3**, 155–174.

Meehl, P. E., & Hathaway, S. R. The K factor as a suppressor variable on the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525–564.

McKinley, J. C., Hathaway, S. R., & Meehl, P. E. The Minnesota Multiphasic Personality Inventory: VI. The K scale. *J. consult. Psychol.*, 1948, **12**, 20–31.

Page, H. A., & Markowitz, Gloria. The relation of defensiveness to rating scale bias. *J. Psychol.*, 1955, **40**, 431–435.

Sarason, I. G. The relationship of anxiety and "lack of defensiveness" to intellectual performance. *J. consult. Psychol.*, 1956, **20**, 220–222.

Sweetland, A., & Quay, H. A note on the *K* scale of the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1953, **17**, 314–316.

Wheeler, W. M., Little, K. B., & Lehner, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, **15**, 134–141.

---

[3] This is reminiscent of our early experience with the Rorschach, when we were astounded at the number of our friends and colleagues who, by clinical criteria, were schizophrenic.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

BRIEF REPORTS

MANIFEST ANXIETY, INTELLIGENCE, AND COLLEGE GRADES[1]

CHARLES D. SPIELBERGER AND WILLIAM G. KATZENMEYER
Duke University

Several studies recently reported in this journal have examined the relationship between the Taylor Manifest Anxiety Scale (MAS) and college grade-point averages (GPAs). Matarazzo et al. and Klugh and Bendix found nonsignificant MAS-GPA correlations of −.08 and .01, respectively; Sarason reported that analysis of variance failed to reveal any changes in GPA as a function of MAS score. Although ACE scores were available to all of these *E*s, only Klugh and Bendig took intelligence into account in evaluating the relationship between MAS scores and grades, and none considered the possibility that anxiety might be differentially related to grades for students with differing intellectual abilities.

The moderately high correlations which have been consistently reported between various measures of intelligence and college grades would seem to indicate that poor academic performance was primarily determined by limited ability while good grades were largely determined by superior intellectual endowment. Thus, personality or motivational variables might be most likely to influence the academic performance of students with average ability. The purpose of the present study was to examine additional evidence on the relationship between MAS scores and GPAs and to determine whether this relationship varied as a function of the intellectual level of the student.

The MAS was administered to all students in the introductory psychology course at Duke University for six consecutive semesters, and GPAs were obtained for that particular semester during which a student had taken the MAS. MAS scores, ACE scores, and GPAs were available for 640 men. The intertest Pearson *r*'s were MAS-GPA, −.14; MAS-ACE, −.11; ACE-GPA, .29 (all significant at .01 level). Even when intelligence was not taken into account, there was a small inverse relationship between MAS scores and grades in the present study. In order to determine whether this relationship was influenced by the intellectual ability of the *S*s, the total sample was divided into the following groups on the basis of ACE scores: lower 20%; middle 60%; and upper 20%. The MAS-GPA correlations for these groups were −.04, −.18, and −.05, respectively. Tests for linear and curvilinear regression indicated GPAs were unrelated to MAS scores for the low and high intelligence groups; the test for linear regression for the middle group yielded an F of 13.06 ($df = 1,390$, $p < .001$). Grades varied inversely with anxiety level for the average aptitude *S*s. College work appeared to be too difficult for low aptitude students whose poor grades were unrelated to their MAS scores. High aptitude students tended to obtain good grades regardless of their anxiety level.

Previous studies may have failed to find a significant relationship between MAS scores and grades for one or more of the following reasons: (*a*) failure to take intelligence into account; (*b*) heterogeneity with respect to intelligence; (*c*) inclusion of both male and female *S*s; (*d*) other factors discussed in the extended report in which these studies are compared.

Brief Report.
Received September 30, 1958.

---

[1] An extended report of this study may be obtained without charge from Charles D. Spielberger, Department of Psychology, Duke University, Durham, North Carolina, or for a fee from the American Documentation Institute. Order Document No. 5889, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

PROGNOSTIC FACTORS IN SCHIZOPHRENIA[1]

BETH STONE[2]
State University of Iowa

In an attempt to identify some of the prognostic factors in schizophrenia, Schofield, Hathaway, Hastings, and Bell (1954) obtained follow-up data on schizophrenic patients and compared good and poor outcome groups in terms of 200 items pertaining to their general history, previous personality, and presenting illness. Using the items which discriminated best, they devised an eight-point prognostic scale which, when applied to the same subjects (*S*s), identified 59% whose posthospital adjustment had been good and 82% whose adjustment had been poor. Since this did not constitute adequate validation, an attempt was made in the present study to cross-validate their scale.

The *S*s were patients admitted to the Iowa State Psychopathic Hospital with a diagnosis of schizophrenia during the years 1929–1933. They were part of a larger group studied by Malamud and Render (1939) who made personal follow-up contacts in 1938. The case histories as well as Malamud's follow-up reports on these patients are still available. Even though *S*s of the present study were hospitalized some 5 to 15 years earlier than Schofield's *S*s, the treatment of both groups was quite similar, neither group having received electroshock, insulin, or metrazol therapy. Both groups were followed up at least five years after their hospital admission.

On the basis of the same five-step scale which Schofield et al. used (1954, p. 157), each *S* was assigned a rating in regard to his posthospital adjustment. Following the procedure of the Minnesota investigators, *S*s who fell into the four extreme outcome categories were used as criterion groups. Our Good outcome group consisted of 19 *S*s, and our Poor group of 64 *S*s. Without the rater having knowledge of the outcome ratings, each patient was rated on the eight-point prognostic scale (Schofield et al., 1954, p. 163). Intra- and interrater reliability checks yielded 91% agreement and 86% agreement respectively.[3]

Since most of the case histories contained no information regarding the patient's school deportment, it was necessary to omit Item 1 of the prognostic scale in the final analysis of the data. In view of this alteration of the scale, the present study is not a complete cross-validation. Application of this seven-point scale to the patients in the Good and Poor groups, using a cutting score of three or more points to identify cases with poor prognosis, resulted in identification of only 42% of the cases in the Good group and 58% in the Poor group—clearly no better than chance identification. It should be noted that 20% of the total sample was in the Good group and 68% in the Poor group. An item analysis revealed that none of the items in the scale discriminated between the two outcome groups at the .05 level of significance.

Brief Report.
Received October 22, 1958.

REFERENCES

MALAMUD, W., & RENDER, N. Course and prognosis in schizophrenia. *Amer. J. Psychiat.*, 1939, **95**, 1039–1057.

SCHOFIELD, W., HATHAWAY, S. R., HASTINGS, D. W., & BELL, D. M. Prognostic factors in schizophrenia. *J. consult. Psychol.*, 1954, **18**, 155–166.

---

[1] An extended report of this study may be obtained without charge from Beth Stone, Dept. of Psychiatry, State Univer. of Iowa, Iowa City, Ia. or for a fee from the American Documentation Institute. To obtain it from the latter source, order Document No. 5890, remitting $1.25 for microfilm or $1.25 for photocopies.

[2] Now a graduate student at Washington University, St. Louis.

[3] Julia S. Weinberg is gratefully acknowledged for making reliability ratings.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

SOCIAL DESIRABILITY IN THE SENTENCE COMPLETION TEST[1]

VITALI V. ROZYNKO

Veterans Administration Hospital, Seattle, Washington

This study attempted to show that: (*a*) the Social Desirability (hereafter called SD) of sentence stems is directly related to the SD of the responses to those stems, and (*b*) the tendency to give socially desirable responses on an inventory type test was related to a similar response tendency on a sentence completion test.

Fifteen sentence stems were selected so that their SD scale values were relatively evenly distributed throughout the rating range. The responses to these stems of 50 psychiatric patients at Seattle VAH were then rated for SD by 50 judges. The mean reliability coefficient for these ratings adjusted by the Spearman-Brown Prophecy formula was .84.

SD scores were computed for each sentence stem and for each patient. These scores were compared with the SD scale values of the sentence stems and with the patient's scores on the SD scale of the MMPI. The SD scale values of the sentence stems correlated .79 ($p = .01$) with the mean SD ratings of the responses to the stems. The patients' sentence completion SD scores correlated .35 ($p = .05$) with their SD scale scores. Five groups of 10 patients each were selected on the basis of the height of the patients' SD scale scores. These groups were then compared with respect to their sentence completion SD scores by an analysis of variance design. The differences between these groups were highly significant ($F = 4.14$, $p = .01$). This high F ratio was primarily due to the mean differences between the high SD scale scorers and the four lower scoring groups. The mean difference between the two highest scoring groups yielded the smallest t, that of 2.93 ($p = .01$). There were no significant differences between any of the four lower scoring groups.

It was concluded that: (*a*) the SD of the sentence stem performs the function of establishing the direction of the response. A socially desirable stem tends to evoke a socially desirable response and vice versa; (*b*) that only patients scoring extremely high on the SD scale tend to perpetuate this response set on the sentence completion test. Meltzoff (1951) reports similar results as to the effect of the sentence stem on responses of college undergraduates. He also found that students placed in stress situations gave a greater number of positively toned responses on a sentence completion test than those who took the test under nonstress conditions. These results suggest that the psychiatric patients in this study who scored high on the SD scale perceived the testing situation as threatening and attempted to hide their "weaknesses" from the hospital staff.

Brief Report.
Received October 25, 1958.

REFERENCE

MELTZOFF, J. The effect of mental set and item structure upon responses to a projective technique. *J. abnorm. soc. Psychol.*, 1951, **46**, 177–189.

---

[1] An extended report of this study may be obtained without charge from Vitali V. Rozynko, VAH American Lake, Wash., or for a fee from the American Documentation Institute. Order Document No. 5891, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

AGREEMENT AMONG RATERS ON THE MULTI-DIMENSIONAL SCALE FOR RATING PSYCHIATRIC PATIENTS[1]

C. JAMES KLETT AND JULIAN J. LASKY
VA Central NP Research Unit, Perry Point, Maryland

The MSRPP (Lorr, 1953) has provided the principal criteria of behavior change in a series of VA cooperative studies of psychopharmacologic treatment of acute and chronic schizophrenic patients. Data were collected as part of one of these studies to provide information on the rater consistency of the scale, and additional data relative to the stability over time of ratings of patients in a placebo group were available from another. The results can serve as a supplement to findings recently reported by Stilson, Mason, Gynther, and Gertz (1958).

A total morbidity score and 11 factor scores are derived from 49 individual scales. Most of these scales are rated from impressions gained during an evaluative interview, and the remainder from ward observations. In the VA studies, each patient was interviewed by a clinical team who later recorded consensus judgments for this part of the scale. A ward team of nurses and psychiatric aides completed the ward items in a similar manner. In one study, the members of each rating team rated patients independently prior to pooling impressions. Of the 563, newly admitted, acute and chronic patients on whom independent ratings were made at a pretreatment evaluation period, 414 were rated by two clinical and two ward team members, and 149 by three clinical and three ward team members. Because the interview and ward ratings must be combined to obtain morbidity and factor scores, ward and interview ratings were paired at random (within each set) to form a scorable set of ratings. Intraclass correlation was used to determine the agreement among raters. For total morbidity this coefficient was .82; for the factors, as follows: A. Retarded depression vs. manic excitement, .72; B. Compliance vs. resistiveness, .65; C. Paranoid projection, .75; D. Activity level, .71; E. Melancholy agitation, .63; F. Perceptual distortion, .80; G. Motor disturbances, .71; H. Submissiveness vs. belligerence, .73; I. Withdrawal, .71; J. Self-depreciation vs. grandiose expansiveness, .72; K. Conceptual disorganization, .80.

Additional ratings were available for 179 mostly chronic schizophrenic males before and after six weeks of inclusion in a lactose placebo group. The product-moment correlation coefficient for total morbidity was .74. Stilson et al. (1958) reported a similar coefficient of .80 with chronic schizophrenic females when three weeks intervened between ratings.

Brief Report.
Received January 2, 1959.

REFERENCES

LORR, M. Multidimensional scale for rating psychiatric patients, hospital form. *U. S. Vet. Adm. tech. Bull.*, 1953, Nov. 16, TB-10-507.

STILSON, D. W., MASON, D. J., GYNTHER, M. D., & GERTZ, B. An evaluation of the comparability and reliabilities of two behavior rating scales for mental patients. *J. consult. Psychol.*, 1958, **22**, 213–216.

---

[1] An extended report of this study may be obtained without charge from the VA Central NP Research Unit, VA Hospital, Perry Point, Md., or for a fee from the American Documentation Institute. Order Document No. 5892, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

VALIDITIES OF WAIS SHORT FORMS IN A PSYCHIATRIC POPULATION[1]

WILLIAM HOWARD
Central Louisiana State Hospital

Using the method advocated by McNemar (1950), Maxwell (1957) has presented the 10 highest short form, full-scale correlations for combinations of two, three, four, and five WAIS subtests. The purpose of the present research was to determine similar coefficients from the scores of psychiatric patients.

*S*s were 186 hospitalized patients who had been administered the WAIS previously by a variety of examiners. Of this total, 70 were schizophrenics, 27 were psychoneurotics, 27 had personality disorders, 22 were manic-depressives, and the remainder were scattered among a variety of diagnostic groups. By education, the *S*s ranged from none to a college degree, and the median was around nine school years.

Short form, full-scale correlations were computed for the same combinations, five each of duads, triads, tetrads, and pentads that Maxwell found to yield the highest coefficients. Analysis was made separately for three age groups.

Age group 25 to 34 produced coefficients which ranged from .929, the lowest duad, to .985, the highest pentad. For age group 35 to 44 the coefficients ranged from .886 to .971. In age group 45 to 54 the range was .913 to .985.

The results for age group 25 to 34 are higher than those found by Maxwell. This discrepancy appeared to be due to the greater variance of the psychiatric full-scale scores, which was significant at the .02 level.

The present author computed short form (same 20 combinations), full-scale correlations for Wechsler's (1955) 45 to 54 age group by means of McNemar's (1950) general formula. The coefficients ranged from .909 to .977, findings which are comparable to those observed in the similar psychiatric group. There was no significant difference in full-scale variance between the two groups at this age.

In addition to reliable short form full-scale correlations, pairs of subtests were found to correlate highly with the verbal and performance scales. None of the coefficients studied fell below .908.

In general, the short form, full-scale correlations were higher for the psychiatric patients than for the standardization sample, but the differences appeared to result from the greater variance of the patient sample. Since the correlations in the two groups were comparable, it was concluded that any of the short forms studied could be used to estimate reliably the intelligence of psychiatric patients.

Brief Report.
Received January 9, 1959.

REFERENCES

McNemar, Q. On abbreviated Wechsler-Bellevue scales. *J. consult. Psychol.*, 1950, **14**, 79–81.

Maxwell, Eileen. Validities of abbreviated WAIS scales. *J. consult. Psychol.*, 1957, **21**, 121–126.

Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corp., 1955.

---

[1] An extended report of this study may be obtained without charge from William Howard, P. O. Box 31, Pineville, Louisiana, or for a fee from the American Documentation Institute. Order Document No. 5893, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 3, 1959

PSYCHOLOGICAL TEST REVIEWS

New Tests

Sanders, Joseph R. *Subsumed Abilities Test*. Age 6–adults. 1 form. 30 min. Test Booklet, with manual, pp. 25, and scoring key and tabulation sheet. New Rochelle, N. Y.: Bruce, 1957.

The *Subsumed Abilities Test* (S.A.T.) is a pencil-and-paper derivative of the Goldstein-Sheerer *Color-Form Sorting Test* and is available only for research use. Like its parent, it requires each subject to make more than one appropriate response to the same visual stimulus. All 30 items, of which the test is comprised, present either four geometrical figures or four pairs. The task is to pick out two sets of three in each item (the two sets may be identical) that are alike in some way. Each of the three abilities—to recognize, to abstract, and to conceptualize—is purportedly tapped by one third of the items. One of the two responses to an item engages the ability at a simple level, the other at a more complex level. The scoring key and tabulation sheet are arranged to obtain a score for each ability and a total score.

The manual offers a rationale for the test that suggests a hierarchy of six abilities: simple recognition, complex recognition, simple abstraction, complex abstraction, simple conceptualization, complex conceptualization. Little data are currently available and less are cited. The manual presents tables of percentile equivalents for samples drawn from grade school, college, professional, secretarial, and "white collar" applicants. Some of these are as small in number as 42.

As the stamp on the Test Manual says, "for research use only."—*E. S. B.*

Jastak, Joseph F. *Jastak Test*. Junior High School. 1 form. Approximately 60 min. (one subtest is administered orally). Test booklet; manual, 58 pp.; scoring keys; and class list. Philadelphia: Educational Test Bureau, 1958.

The *Jastak Test* represents a bold attempt to combine in one measuring process indices of potential ability and personality. Ten subtests, most of them familiar versions of items used in other intelligence tests and with time limits ranging between 3 and 8 minutes, provide the basic data. Each subtest score can be converted to a standard score and combined to obtain a conventional intelligence index. The remaining scores are derived from various aspects of the scatter of subtests. Thus, there is a capacity score, which, by an incompletely described process, uses the individual's highest subtest performance as a basic for estimating his capacity as distinguished from his current level of functioning. The intra-individual variance in the subtests is further utilized in various combinations to derive personality factors identified as language, reality, motivational, and psychomotor. The format is attractive and the test items appear designed to maintain interest. There seems to be a good balance of intellectual factors sampled.

Considering the fact that teachers and administrators are invited to use this test, the manual has several defects. Much of the technical test development discussion is too complex to have much meaning to this audience and is too brief for the technically trained reader. No references to fuller descriptions of the process of analysis are cited. Samples of adequate size provide the basis for norms, reliability indices, and intercorrelational and factor studies. The reliabilities of the subtests, with only few exceptions, range between .80 and .88. The intercorrelations between the range scores which form the basis for each of the four personality scores are disquietingly low. Though the manual contains many warnings about pitfalls in the administration of the test and the interpretation of the results, there are no warnings about the experience and training necessary to an evaluation of a child's contact with reality or other symptoms of maladaption.

When one considers that this instrument is claimed to measure both intelligence and personality, it is surprising to find the manual completely bare of validation data. Reference is made to the author's clinical experience, and several case histories are presented. This is weak evidence to marshal against the many discouraging results of other efforts to demonstrate the value of scatter in intelligence subtests as indices of personality. Not even evidence of the test's ability to predict school performance is offered. Its use for this latter purpose probably can be supported on the basis of the precedence provided by other tests of similar content that have been demonstrated to have similar value. Without validation data, the *Jastak Test* cannot be accepted as a way of obtaining the other inferences it purports to provide—*E. S. B.*

Ford, J. S., Grimsley, G., Ruch, F. L., & Warren, N. D. *Employee Aptitude Survey.* College and adult. 2 forms. 55 min. 10 separate subtest booklets, Form A or Form B ($2.50 per 25, $6.00 per 100, and $25.00 per 500 for each subtest, includes manual and scoring stencils); a general manual, 6 pp.; a manual, 4 pp., and scoring stencil for each subtest; a manual for interpretation, 11 pp. Los Angeles: Psychological Services, 1956.

With few exceptions the 10 subtests in this battery (verbal comprehension, numerical ability, visual pursuit, visual speed and accuracy, space visualization, numerical reasoning, verbal reasoning, word fluency, manual speed and accuracy, and symbolic reasoning) are familiar versions of other tests. The relatively high reliabilities (range .7 to .8) obtained for such brief tests suggest a careful item analysis. Because the items are incorporated into the answer sheet, the distribution of test materials can be expedited.

There are several respects in which the manuals for these tests are lacking. In a few instances no data on the correlation between alternate forms is offered. Information on the intercorrelation of the subtests is given in a form of coefficients of independence which conceal as much as they reveal. The unwary reader will overlook the fact that this coefficient $[100\,(1 - r^2)]$ not only reflects the variance specific to the test but also the variance due to errors of measurement. In the case of symbolic reasoning, its coefficient of independence with respect to verbal reasoning is tabled as .64 for a sample of 90 high school students. Working backward, it becomes apparent that the correlation between the two tests is .60, a high degree of overlap when one considers that their respective test-retest reliabilities are .69 and .79. An intercorrelation table is a franker and more direct way to communicate the necessary information.

With the exception of word fluency, a considerable amount of impressive validation evidence is offered for each test. The reported multiple correlation of .83 with one-year grade-point averages for a sample of engineering students would be considered fantastic were it not topped by the report that a cross-validation of the regression equation in the next year's sample against a pass–fail criterion produced a shrinkage of only .03. Unfortunately, no citations of published studies are offered, and only very brief statements of validity studies are given.

If the validity studies cited are to be credited, the *Employee Aptitude Survey* may be one of the most efficient test batteries we have for use with the wide range of subjects represented by high school and college students, and by adult workers.—*E. S. B.*

PSYCHOLOGY AND REHABILITATION

BEATRICE A. WRIGHT, Ph.D.
Editor

Based on the proceedings of an Institute on the roles of psychology and psychologists in rehabilitation, sponsored by the American Psychological Association and supported by a grant from the Office of Vocational Rehabilitation, Department of Health, Education and Welfare.

The volume examines the principles and assumptions underlying the modern philosophy of rehabilitation, paying special attention to the roles and functions of psychologists as members of an interdisciplinary team. Problems of training are explored in detail and in the context of the changing nature of the field. Areas of needed research are highlighted.

Price $1.50

Order from:

American Psychological Association, Inc.
Dept. PAR
1333 Sixteenth Street, N.W.
Washington 6, D.C.

Outstanding McGRAW-HILL *Books . . .*

CLINICAL PSYCHOLOGY—The Study of Persons

By Richard W. Wallen, Creelman Associates. *McGraw-Hill Series in Psychology.* 400 pages, $6.50

The major methods used by clinical psychologists are presented to help students prepare for actual practice. The treatment of the interview, projective methods, psychometric tests, and psychotherapy not only gives a survey of the clinician's work, but establishes a foundation for further, specialized courses. Case material is given to help the student become more perceptive and more able to think like professional clinicians. An instructor's manual is available for those using the text.

NONPARAMETRIC STATISTICS:
For the Behavioral Sciences

By Sidney Siegel. Pennsylvania State University. 330 pages, $6.75.

The first book-length treatment of nonparametric, or distribution-free statistics. It covers comprehensively the nonparametric statistical tests and measures of correlation, demonstrating their usefulness in behavioral science research. It is written for the reader with no special training in mathematics, and is organized to serve as a reference work and text.

SCIENTIFIC METHOD IN PSYCHOLOGY

By Clarence W. Brown and Edwin E. Ghiselli, University of California, Berkeley. *McGraw-Hill Series in Psychology.* 380 pages, $6.50.

This timely work describes the method of science as applied to the study of psychological phenomena. The text considers first the underlying logic of science as it relates to the study of human behavior, and then describes the commonly accepted steps used in conducting a scientific study of a psychological phenomenon. In the latter part of the book specific procedures and applications of the scientific methods in psychology are described.

PERSONALITY

By J. P. Guilford, University of Southern California. *McGraw-Hill Series in Psychology.* 584 pages, $7.50.

A text for upper-division psychology and education students which places emphasis on individual differences, rather than on stimulus-response psychology of individual behavior. The approach is from the standpoint of the logic of experimental method. Many of the newer methods of personality assessment are described. The author develops a model of personality based upon factor theory as a basis for the assessment and description of persons. Included are chapters describing personality in terms of dimensions found by factor analysis, including dimensions of morphology, physiology, aptitude, and pathology, as well as of temperament, needs, interests, and aptitudes.

Send for copies on approval

McGRAW-HILL BOOK CO., INC.

330 West 42nd Street, New York 36, N. Y.

JOURNAL OF CONSULTING PSYCHOLOGY

August, 1959 Vol. 23, No. 4

Contents

ARTICLES

The Factorial Structure of the WISC at Ages 7–6, 10–6, and 13–6: *Jacob Cohen* - - - 285

Direct Measurement of Cognitive Deficit in Brain-Injured Patients: *Harold L. Williams, Ardie Lubin, and Charles F. Gieseking* - - - 300

The Inhibition Process, Rorschach Human Movement Responses, and Intelligence: Some Further Data: *Murray Levine, George Spivack, and Byron Wight* - - - 306

Prognosis of Social Adjustment for Mental Hospital Patients Under Psychotherapy: *Margaret J. Rioch and Ardie Lubin* - - - 313

MMPI *L-F-K* Relationships With Criteria of Behavioral Disturbance and Social Adjustment in a Schizophrenic Population: *Leonard R. Gross* - - - 319

The Interaction Among Anxiety, Stress Instructions, and Difficulty: *Gerald Wiener* - - 324

Relationships Among Direct and Indirect Measures of the Achievement Motive and Overt Behavior: *David Marlowe* - - - 329

An Appraisal of Taulbee and Sisson's "Configurational Analysis of MMPI Profiles of Psychiatric Groups": *Sol L. Garfield and Jon Sineps* - - - 333

The Effects of Warm and Cold Interaction on the Administration and Scoring of an Intelligence Test: *Joseph Masling* - - - 336

More Construct Validation of the Ego-Strength Scale: *Irving I. Gottesman* - - - 342

Validation of a Need Scaling Technique for the Adjective Check List: *Alfred B. Heilbrun, Jr.* - - - 347

The Screening Value of the Cornell Medical Index: *M. Powell Lawton* - - - 352

Symbolism Validity and Learning Without Awareness: *Caroline Taylor MacBrayer* - - 357

The Social Desirability of Trait Descriptive Terms: Applications to a Self-Concept Inventory: *Emory L. Cowen and Phoebus N. Tongas* - - - 361

Two Questions: A Reply to Cowen and Tongas: *Robert E. Bills* - - - 366

A Contribution to the "Meaning" of Rorschach's Inkblots via the Semantic Differential: *A. I. Rabin* - - - 368

What Does the Rorschach *Z* Score Reflect?: *Leon S. Otis* - - - 373

BRIEF REPORTS

Ecologic Factors in the WAIS Picture Completion Test: *Bernard L. Bloom* - - - 375

Personality Implications of Cigarette Smoking Among College Students: *Daniel S. P. Schubert* - - - 376

"Social Desirability" and "Anxiety" Variables in the IPAT Anxiety Scale: *A. W. Bendig* - - - 377

Sex Differences in Body Concepts: *George Calden, Richard M. Lundy, and Richard J. Schlafer* - - - 378

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual*. Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

> [1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual*.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

THE FACTORIAL STRUCTURE OF THE WISC AT AGES 7–6, 10–6, AND 13–6[1]

JACOB COHEN

Franklin D. Roosevelt Veterans Administration Hospital and
New York University

The present article continues a systematic investigation of the intellectual domain as sampled by the Wechsler scales (Wechsler: 1949, 1955, 1958). Previous articles have described the factorial structure of the Wechsler-Bellevue (W-B) for male veteran neuropsychiatric patients (Cohen: 1952a, 1952b) and of the Wechsler Adult Intelligence Scale (WAIS) for standardization samples of adults over a wide age range (Cohen: 1957a, 1957b). These and other factor-analytic studies are summarized by Wechsler (1958, Ch. 8). The present study is concerned with the data provided by the standardization of the Wechsler Intelligence Scale for Children (WISC), a test of form and content similar to those above, on standardization samples of children at age levels 7–6, 10–6, and 13–6.

This study, then, provides an extension downward of the age range to age 7–6. Taken by itself, it provides some insight into the process of intellectual maturation via the comparative analysis of the factorial structures for the three age groups, and the accompanying by-product of a rationale for the subtests. By comparing its results with those of the previous studies, it makes possible the presentation of a broad picture of intellectual organization on Wechsler scales over widely varying conditions of age and psychopathology.

Method

Subjects

The data analyzed were those of the three age levels of the total standardization samples reported in the manual (Wechsler, 1949). The standardization sample was stratified at each age by geographic area, urban-rural residence, and parental occupation to provide a fairly close match with the child population distribution of the 1940 United States census. The *N* of each age group was 200, made up equally of boys and girls, all of whom were within 1½ months of their midyear (Wechsler, 1949, pp. 7–8).

Analysis

The matrices of intercorrelations of the three age groups given in the manual (Wechsler, 1949, pp. 10–12) were separately subjected to the following analysis:[2]

1. Thurstone's complete centroid method was used (Thurstone, 1947, pp. 161–170), with communalities estimated by his Equation 15 (Thurstone, 1947, pp. 300, 318). Since the absolute discrepancies between estimated and computed communalities were small (median = .02), the factorizations were not reiterated. The decision as to the number of factors to be accepted for rotation was based on the statistical criteria of Saunders (Cattell, 1952, pp. 300–301), McNemar (1942a) and Burt (Cureton, 1955), and on the results of the previous factorization of the WAIS (Cohen, 1957b). The statistical criteria were inconsistent with regard to the acceptance of the fifth centroid, but since the WAIS had yielded five factors in three adult age groups, and since the extraction of residual (i.e., "error") factors results in their vanishing in

---

[1] From the Psychiatric Evaluation Project of the Psychology Service, Veterans Administration Hospital, Montrose, New York. The author gratefully acknowledges the cooperation of the management of the hospital, and is especially indebted to Catherine S. Henderson for typing the manuscript.

[2] To save printing costs, tables giving centroid loadings and communalities, transformation matrices and intercorrelations among primaries have been deposited with the American Documentation Institute. Order Document No. 5923, remitting $1.25 for microfilm or $1.25 for photocopies.

the process of rotation, five centroids were extracted in each group. The "reality" of the fifth factor is attested to by its consistency among the three children's groups and with the adult groups on the WAIS (Cohen, 1957b).

2. Rotation was essentially blind and independent in the three groups, and was performed by Thurstone's method of two-dimensional sections (Thurstone, 1947, pp. 194–216). The rotation criteria were oblique simple structure and a positive manifold with maximization of the number of variables in a ±.05 and ±.10 hyperplane (Cattell, 1952, pp. 235–236).

3. The intercorrelations among the oblique primary factors were determined and subjected to a second-order general factor analysis (Thurstone, pp. 273–277, 421–434), as a product of which the correlations of the subtests with this general factor were obtained. This analysis made possible the determination of the proportion of total and of true (nonerror) variance attributable to the second-order general factor and comparisons with these proportions found for adults on the WAIS (Cohen, 1957b).

Results and Discussion

Five oblique primary factors were found in each age group. Factor loadings in excess of .20 are accepted as significant for the purpose of factor interpretation.[3] To facilitate comparisons between age groups in Table 1, loadings between .20 and .39 are followed by an asterisk (*), and loadings of .40 and higher by a double asterisk (**). The factors were found to be very similar in the three age groups, a finding to which we have grown accustomed (Cohen: 1952b, 1957b). Table 1 is accordingly organized by factor. Decimal points are omitted in the factor tables throughout.

Factor A: Verbal Comprehension I

The subtests do not load this factor in the three groups of children with quite the neat consistency which obtained for the four adult groups on the WAIS (Cohen, 1957b, p. 284). All three children's groups load Information and Similarities, Arithmetic fails to load in the oldest (13–6 years old) group, Comprehension loads only in that group, while Vocabulary loads in the youngest and oldest groups—in the latter heavily (see Table 1-A). These tests are all verbal, and the factor is named Verbal Comprehension I, the roman numeral being used to distinguish it from Factor D, which also involves verbal subtests.

Since there are only five verbal subtests in all, no confident interpretation of Factor A which distinguishes it from Factor D (named Verbal Comprehension II) can be made. However, an hypothesis consistent with the observed loadings on the two factors can be offered. Factor A seems to reflect that aspect of verbally retained knowledge impressed by formal education: facts (Information) and verbal categorizing (Similarities) at all three ages, and number manipulation (Arithmetic) at ages 7–6 and 10–6. The pivotal subtest is Information, which does not appear on Factor D (or any other). Factor D, on the other hand, seems to stress judgment and is discussed below. The differentiation between the two factors should not be overstressed in the light of the correlations between them of .78, .82, and .43 in ascending order of age in the three groups. (Indeed, when factor scores come to be discussed below, it will be seen that these two factors are combined into a single verbal factor.)

By age 13–6, the composition of Factor A is "stabilized" in that it is loaded by the same four subtests as appear on it throughout the adult years on the WAIS: Information, Comprehension, Similarities, and Vocabulary (Cohen, 1957b). Factor A, then, is identified as the well-established verbal comprehension factor found in all previous factor-analytic studies of the W-B and WAIS with normals at different age levels (cf. Balinsky, 1941;

---

[3] The use of the .20 significance criterion in place of the more conventional (and arbitrary) .30 is justified as follows: This investigation can be viewed as a set of three replications of a factor analysis (with the WAIS study [Cohen, 1957b] providing four more); therefore, consistency between groups is an effective guard against the acceptance of nonsignificant loadings as significant, and superior to the rule-of-thumb .30 criterion, which would lead to spuriously inconsistent results between groups. Further justification of this criterion is provided by the fact that of the 180 factor loadings, only six were between .10, the upper limit of the hyperplane, and .20.

Table 1

Oblique Primary Factors for the Three Age Groups

		A Verbal Comp. I			B Percept. Org'n.			C Free. Distract.			D Verbal Comp. II			E (Quasi-Specific)		
Subtest	Age:	7–6	10–6	13–6	7–6	10–6	13–6	7–6	10–6	13–6	7–6	10–6	13–6	7–6	10–6	13–6
Inf.		41**	32*	39*	−02	02	07	−09	−02	10	02	15	00	17	00	−10
Comp.		−10	−06	37*	−02	−01	−10	−09	06	−07	47**	38*	20*	09	−05	07
Arith.		21*	41**	07	−04	04	−04	17	10	36*	−06	−09	02	18	08	−05
Simil.		34*	38*	26*	09	−05	02	09	02	00	−01	08	24*	−10	−06	09
Vocab.		21*	05	47**	−07	07	00	−04	07	−10	26*	30*	−01	05	00	08
D. Sp.		−04	04	02	−04	00	01	38*	44**	33*	02	−09	00	04	−04	−10
P. C.		−06	04	−05	06	43**	50**	09	−10	04	28*	21*	27*	−10	−04	−10
P. A.		01	−02	−02	29*	07	09	33*	−09	−06	−01	09	14	−10	38*	33*
B. D.		06	02	07	50**	55**	46**	−08	15	01	07	03	10	09	04	−03
O. A.		−01	−05	−03	59**	49**	57**	07	32*	−09	−05	−05	−08	09	−04	10
Coding		−07	03	01	06	−05	−01	06	09	10	−01	−03	−10	39*	23*	25*
Mazes		05	−02	−05	56**	42**	03	−04	20*	23*	−02	02	−03	04	01	10

Birren, 1952; Cohen, 1957b; Wechsler, 1958) as well as with psychiatric patients (cf. Cohen, 1952a; Cohen, 1952b; Hover, 1950; Simkin, 1950), which in turn is the same factor found in all general investigations of the intellectual domain (cf. Thomson, 1951).

Factor B: Perceptual Organization

This factor is readily identified with Factor B in the adult group. The same two subtests, Block Design and Object Assembly, load it consistently and exclusively in the three children's age groups (see Table 1-B) as they did in the four adult age groups (Cohen, 1957b). The affinity to this factor of Picture Arrangement seen here at age 7–6 is also found in some adult groups (Cohen, 1957b, p. 285). Picture Completion loads here at ages 10–6 and 13–6, and Mazes, which has no match on either the WAIS or W-B, appears at ages 7–6 and 10–6.

This factor has appeared consistently in previous factor-analytic studies of Wechsler scales, and although it has been variously named, there is no mistaking its identity (see references in Cohen, 1957b, and Wechsler, 1958). The tests which load it are all nonverbal and require the interpretation and/or organization of visually perceived materials against a time limit. It is thus given the same name as in the WAIS study, Perceptual Organization (Cohen, 1957b).

Factor C: Freedom from Distractibility

In Table 1-C, the consistently and exclusively loading subtest is Digit Span, while Mazes loads at 10–6 and 13–6, and Picture Arrangement, Object Assembly, and Arithmetic each load in a different group. At 13–6, the two significantly loading subtests, Arithmetic and Digit Span, are the same ones as appear throughout the adult range (Cohen, 1957b).

This factor has appeared repeatedly in previous factorial studies of the Wechsler subtests, and has been variously interpreted as memory (Balinsky, 1941; Birren, 1952; Cohen: 1957a, 1957b), freedom from distractibility (Cohen: 1952a, 1952b), attention-concentration (Hover, 1950), and concentration-speed (Simkin, 1950). In the WAIS study (Cohen, 1957b), it was pointed out that these concepts are not as diverse as they may seem at first glance, since rote memory requires as a precondition the ability to remain undistracted (to attend or concentrate), and C was interpreted to be a memory factor. In the present study, it is felt necessary to revert to the original interpretation of this factor as one of freedom from distractibility (Cohen: 1952a, 1952b) primarily due to the loadings of subtests which clearly do not involve memory (Mazes, Picture Arrangement, Object Assembly), but which it seems reasonable to suppose are quite vulnerable to the effects of distractibility. To make the point explicit, Factor C in Wechsler scales (WISC, WAIS, or W-B) is primarily a Freedom from Distractibility factor; its interpretation as a memory factor in the WAIS study (Cohen, 1957b) and others (Balinsky, 1941; Birren, 1952) is in error.

Factor D: Verbal Comprehension II

Comprehension and Picture Completion load at all three age levels, Vocabulary at 7–6 and 10–6, and Similarities at 13–6 (see Table 1-D). This factor seems to change in a systematic way with increase in age: Vocabulary drops out by age 13–6, Comprehension drops from the highest loading in the table at age 7–6 to a barely acceptable one at age 13–6, and Similarities appears weakly at age 13–6; only Picture Completion loads at the same (weak) level throughout. The identification of this factor with Factor D in the adult study is readily made by the apparent continuation of these subtest trends for the 18–19 age group on the WAIS (Cohen, 1957b, p. 286): Picture Completion continues to load (and at the same level), the diminishing Comprehension subtest drops to a zero loading, and Similarities continues its weak loading.

When Factor D in all four adult groups was considered (Cohen, 1957b), it was seen to be almost solely a matter of Picture Completion: only this subtest loaded in all four groups, and the other three subtests which ever load the factor do so in three different groups and with barely acceptable loadings (.20 to .22). The factor was therefore accepted as a quasi-specific in the WAIS study and not interpreted. When the results of the present investigation are added for consideration, a tentative hypothesis about this factor suggests itself, namely, that it reflects the application of judgment to situations following some implicit verbal manipulation. The suggested differentiation from Factor A is that the latter represents formally learned verbal comprehension, while Factor D is the application of verbal skills to new situations. This differentiation is consonant with the findings; e.g., Comprehension declines as a measure of judgment as formally learned "right" answers are acquired.

Since the distinction between these two factors is not sharp, either in conception or in the light of the already noted substantial correlations between them, and since such distinction which has been made is based on *ad hoc* reasoning, Factor D is called Verbal Comprehension II, and neither factor is more specifically identified. This more conservative position simply holds that two highly correlated factors were found loaded primarily by overlapping verbal subtests.

Factor E

As was the case in the adult study (Cohen, 1957b, pp. 286–287), the only consistently loading subtest is Coding, which is equivalent to Digit Symbol (see Table 1-E). Picture Arrangement loads this factor more substantially than does Coding in the older two groups, but fails to load it at all at age 7–6. In the adults one also finds some affinity for Factor E in Picture Arrangement (ages 45–54), but again the only consistently loading subtest is Digit Symbol.

No psychological interpretation of Factor E is offered here, nor was any offered in the WAIS study. It is not an important factor in terms of the amount of variance it accounts for in any of the children or adult groups. But it is "real" in the sense that its consistent appearance in both studies precludes its attribution to sampling fluctuations in the subtest intercorrelations.

Second-Order General Factor

As already noted, the rotations were oblique, and the result was a substantial degree of correlation among the primary factors in all three

Table 2

Correlations With G of WISC Subtests and Primary Factors for the Three Age Groups

Subtest	Age 7–6	Age 10–6	Age 13–6
Information	61	79	78
Comprehension	59	74	63
Arithmetic	61	74	64
Similarities	58	67	72
Vocabulary	69	86	79
Digit Span	58	52	46
Picture Completion	47	42	41
Picture Arrangement	60	67	53
Block Design	48	58	58
Object Assembly	47	43	48
Coding	38	49	55
Mazes	37	49	49
Primary Factor			
A	83	85	91
B	53	29	64
C	87	82	90
D	88	93	33
E	69	91	87

groups.[4] When these matrices of factor intercorrelations are factored in turn, they are found to be substantially (but not completely) accounted for by a single (second-order) general factor. Table 2 gives the correlation of the subtests and the primary factors with this factor, G, which is interpreted as present general intellectual ability. The G correlations of the subtests are moderate (median over all age groups = .58) and of the primary factors quite high (over-all median .84). The degree of correlation with G of the subtests is not quite so high for children on the WISC as it was found to be for adults on the WAIS (for further discussion on this point, see below); still, the magnitudes are substantial and suggest that the WISC Full Scale IQ, which is a linear function of an equally weighted sum of subtest weighted scores, is loaded quite strongly with and, hence, is a good measure of G. This is, in fact, the case. When the correlation of Full Scale IQ with G is determined, using the correlation of sums formula, it is found to be .92, .94, and .95 for the three groups in order of increasing age. The "Verbal IQ" is an equally good measure of G (.93, .94, and .94), while the "Performance IQ" is a relatively poor one (.78, .82, and .81).

As has been found in the past work with the W-B (Cohen: 1952a, 1952b; Wechsler, 1958) and the WAIS (Cohen: 1957a, 1957b; Wechsler, 1958), the "essentially verbal" tests, particularly Vocabulary and Information, are consistently the best measures of G over the age range considered; the poorest are Picture Completion, Mazes, and Object Assembly. Among the primary factors, except for the anomalously low correlation with D at ages 13–6, it is Factor B which is consistently the poorest measure of G, while the others are, on the average, about equally and highly related. These results parallel the G correlations of Verbal and Performance IQs noted above.

In the WAIS study, noteworthy shifts in the correlations of subtests and major factors with G occurred as a function of age in the 60–75+ age group (Cohen: 1957a, 1957b). Over the age range from 7–6 to 13–6 of the present study, no such shifts occur; the subtests and factors maintain fairly consistent levels of correlation in the three children's age groups as they did for the three younger adult age groups, i.e., 18–19, 25–34, and 45–54.

Subtest Specificity

A body of doctrine has come down in the clinical use of Wechsler scales, which involves a rationale in which specific intellective and psychodynamic trait-measurement functions are assigned to each of the subtests (e.g., Rapaport, 1945). This has been used to provide a theoretical basis for both research in pattern analysis and in clinical psychodiagnosis. Implicit in this rationale lies the assumption that a substantial part of a subtest's variance is associated with these specific measurement functions. In previous work with the W-B for neuropsychiatric patients (Cohen: 1952a, 1952b) and with the WAIS for the standardization data (Cohen: 1957a,

---

[4] The median and range of primary factor intercorrelations for the three groups were as follows: Age 7–6, Mdn = .56, $R = .39$–.78; age 10–6, Mdn = .71, $R = .18$–.83; age 13–6, Mdn = .56, $R = .10$–.83. All primary factor intercorrelations are given in the ADI material (see Footnote 2).

1957b), the basis for this doctrine was challenged on grounds of small specific variance for the subtests. When the communality of a subtest (proportion of variance shared with other subtests) is subtracted from its internal consistency reliability coefficient (proportion of true, i.e., nonerror variance), the remainder constitutes the subtest's specificity, the proportion of its variance which is both nonerror and distinctive to it. This component of the subtest's variance, and only this component, can carry the variance necessitated by an hypothesized one-to-one correspondence of subtest to trait in any clinical rationale.

These values for the age groups on the WISC are presented in the Subtest-Specificity section of Table 4. The over-all median of these values is .18, and excepting only Mazes (the only subtest not represented in the adult scales), no subtest at any age has as much as one-third of its variance attributable to specificity. It is apparent that these specificities are quite inadequate to serve as a basis for a subtest-specific rationale.

Thus, it is true for the WISC (as it is for the WAIS in normal adults and the W-B in patients) that a subtest's measurement function is most meaningfully and completely described in terms of G and the primary abilities, which account for the bulk of its reliable variance, and not in terms of its small (essentially uninterpretable and probably complex) specificity. Adherents of the "clinical" rationales can find no support in the factor-analytic studies of Wechsler scales. A detailed rationale for the subtests, based on the present factor analyses, is presented in a later section.

Age, G, and Intellectual Organization

The availability of complete analyses of the same set of intelligence subtests for representative subjects over a wide age range makes possible for the first time direct scrutiny of changes in intellectual organization concomitant with age. In the following discussion, let it be understood that generalizations about intellectual organization carry along the qualification "insofar as Wechsler subtests are broadly representative of the complex of functions considered relevant to intelligence." Despite this qualification, it seems to the author unlikely that another collection of subtests drawn from within the same framework would lead to substantially different conclusions.

It has already been noted that the same correlated primary factors appear at each of the age levels 7–6, 10–6, and 13–6 of the WISC standardization samples, that these are substantially correlated, and thereby give rise to G. The analysis which was performed makes possible the determination of the amount and proportion of variance attributable to each of various identifiable sources:

1. G—the general second-order factor, i.e., present general intellectual functioning; that part of the variance shared by all the subtests.
2. c—the communality (h^2), i.e., that part of the variance common to two or more of the subtests. This includes all influences leading to intercorrelations among subtests (including G).
3. s—that part of the variance which is specific to a given subtest in a battery and consistently measured by it (see preceding section, *Subtest Specificity*).
4. e—random error of measurement variance, i.e., the subtest's "unreliability."

Table 3 represents the average proportions of the subtests attributable to the above sources at each age for the three children's groups on the WISC, as well as comparable

Table 3

Proportions of Total Variance Attributable to Various Sources for the Three Children's Age Groups on the WISC and the Three Adult Age Groups on the WAIS[a]

Age Group	Source G	 c	 s	 e
Children				
7½	30	48	19	33
10½	40	57	18	25
13½	36	54	19	26
Mean	35	53	18	28
Adults				
18–19	53	67	14	19
25–34	50	64	16	20
45–54	54	67	15	18
Mean	52	66	15	19

[a] Rows do not sum to unity because of overlap—G is part of c. However, c, s, and e are mutually exclusive and exhaustive, i.e., these three sources do sum to unity in each row (within rounding errors).

data on three adult groups on the WAIS (Cohen, 1957b). G accounts for .30–.40 of the total variance in the children's groups, the mean over the three groups being .35, and no monotonic age trend is noted.

Inferences about the extent of generality of intellectual functioning cannot be restricted solely to a consideration of G, since some subjectivity occurs in the course of rotating to oblique simple structure. When, alternatively, one scrutinizes the communality (c, Table 3), which is determined objectively and fairly directly from the subtest intercorrelations, a picture similar to that of G emerges for the three children's age groups. Communal variance on the average accounts for about half of the total (.53), with neither large differences nor a monotonic trend with age in the children's groups.

With respect to specificity and error, the former remains constant, while the latter is high in the 7½-year-olds as compared with the two older groups. This simply reflects the fact that the reliability coefficients of the subtests are lower at this age, so that measurement error accounts for about one-third of the variance in contrast with one-fourth for the older children.

Summarizing for the three groups of children, about half of the total subtest variance is shared in common and two-thirds of this communality (.35 of the total) is a function of G. Only .18 of the total variance reflects subtest specificity, while a considerably higher proportion (.28) lies in errors of measurement.

The picture of pregeriatric adult intellectual organization on the WAIS is strikingly different (lower half of Table 3). We can put aside immediately the differences among the three adult age groups—they are very small and can safely be ignored. Turning, then, to a consideration of the mean values, we find that G accounts for fully .52 of the total variance, an amount half again as large as that for the children.[5] For the adults, about two-thirds (.66) of the total subtest variance is shared among the subtests in contrast to the .53 figure for the children. G variance accounts for four-fifths ($.52/.66 = .79$) of the communality for the adults and for only two-thirds ($.35/.53 = .66$) of the communality for the children. The amount of specificity is slightly smaller for adults, and the amount of error variance substantially smaller.

The conclusion that adults show both higher generality and communality of intellectual functioning than children suggests itself with some force. This conclusion is persuasive even in the absence of sampling error formulae in the light of the gap of .10 between the highest G proportion in the children's groups (.40 at age 10–6) and the lowest G proportion in the adult groups (.50 at age 25–34).[6]

This conclusion is in flat disagreement with what is generally held to be the case in the literature. The doctrine of intellectual differentiation with age is most explicitly stated by Garrett (1946): "Abstract or symbol intelligence changes in its organization as age increases from a fairly unified and general ability to a loosely organized group of abilities or factors" (p. 373). Operationally, this would be shown by a decrease between childhood and adulthood of general factor variance, communality variance, mean intercorrelation, or first centroid factor variance.

Garrett (1946), Anastasi (1948), and Burt (1954) review the evidence for this differentiation hypothesis. It is based on some studies which show a reduction in G with age during childhood (although the most extensive of such studies, that of McNemar on the Binet [McNemar, 1942b], does not), and on demonstrations of more G in elementary school children than in college students. The latter is readily explained as due to the selection for intelligence at the college level (which, because of the narrowing of the "range of

---

[5] Note that the statements which characterize adult intellectual functioning refer to adult ages up to age 54. The substantially greater generality compared with children no longer holds for the 60–75+ age group. In this geriatric group, G accounts for only .42 of the total WAIS variance, an amount only slightly greater than that of the children (Cohen, 1957b, p. 289).

[6] These conclusions are not changed when the "true" (i.e., nonerror) variance is analyzed to allow for the lower subtest reliabilities of the children: on the average, (*a*) .49 of the true variance of children's scores is associated with G compared to .64 for adults; (*b*) the communality constitutes .74 of the true variance for children and .82 for adults; (*c*) complementarily, there is substantially more specificity in intellectual functioning in children (.26) than in adults (.18).

Table 4

Proportions of Total Variance Attributable to Various Sources for the WISC Subtests and IQs at Ages 7–6, 10–6, and 13–6

	Source											
	General			Primary-Specificity			Subtest-Specificity			Error		
Subtest Age:	7–6	10–6	13–6	7–6	10–6	13–6	7–6	10–6	13–6	7–6	10–6	13–6
Inf.	37	62	61	23	13	15	06	05	06	34	20	18
Comp.	35	55	40	16	13	18	08	05	13	41	27	29
Arith.	37	55	41	12	11	11	14	18	25	37	16	23
Simil.	34	45	52	10	18	15	22	18	12	34	19	21
Voc.	48	74	62	11	07	11	18	10	17	23	09	10
D. Sp.	34	27	21	11	18	10	15	14	19	40	41	50
P. C.	22	18	17	11	24	34	26	24	17	41	34	32
P. A.	36	45	28	16	10	15	20	16	29	28	29	28
B. D.	23	34	34	28	34	26	33	19	28	16	13	12
O. A.	22	18	23	37	32	48	04	13	00	37	37	29
Coding	14	24	30	17	04	04	29	32	26	40	40[a]	40[a]
Mazes	14	24	24	31	19	04	34	38	47	21	19	25
Mean	30	40	36	19	17	18	19	18	19	33	25	26
							Battery-Specificity					
Verbal IQ	86	89	89				03	05	04	11	06	07
Perf. IQ	61	68	65				27	22	25	12	10	10
F. S. IQ	85	88	91				08	07	04	07	05	05

Note.—Decimal points omitted.
[a] Estimated. (See Wechsler, 1949, pp. 13–14.)

talent," results in low test intercorrelations) rather than to age as such. When unselected children are compared with *unselected* adults in generality of intellectual functioning on representative tests, the differentiation hypothesis is rejected; in fact, we find greater generality of intellectual functioning in adults than in children.

A Rationale for the Subtests

In previous articles, rationales for Wechsler subtests were presented based on factor analyses of the W-B in neuropsychiatric patients (Cohen, 1952a) and of the WAIS standardization population (Cohen, 1957a). In order to adduce a similar rationale for these subtests on the WISC, for each subtest in each age group the *proportion* of variance attributable to the following sources was determined (see Table 4):

1. General: Variance due to G is on the whole larger than the other sources; this simply underscores the importance of the general factor in the WISC. It is of importance that the reader note that the table entries are proportions of variance, not correlations; the latter are higher—the square root of the tabled value—and have been given in Table 2.

2. Primary-Specific: As used here, this concept is meant to identify the proportion of a subtest's variance which is shared with at least one other subtest, *over and beyond G*. In previous analyses (Cohen: 1952a, 1957a), this has been identified as primary-specific variance, i.e., that part of each primary factor's variance specific to the factor and independent of G. The tabled entries here combine the variance of all primary-specifics, although for most of the subtests one factor predominates. The nature of the primary-specific variance of each subtest was determined in the basic analysis (Table 1).

3. Subtest-Specific: This represents what each subtest measures exclusively (and validly) in the scale.

4. Error: This is the proportion of variance accounted for by random error of measurement, i.e., unreliability. These values are, on the average, four-fifths as large as the G proportions, and suggest the wisdom of caution in the interpretation of single subtest scores.

The rationale which follows is for the most part an interpretation of Table 4, to which the reader should freely refer without explicit reference.

Information

This subtest occupies an uncontested second place, after Vocabulary, as a measure of G among the WISC subtests, particularly for older children, where G accounts for three-fifths of its variance. In the 7½-year-olds, it does not perform as well in measuring G, but then at this age level, because of the relatively high unreliability (an average over the subtests of a third of the variance), most of the subtests show less G variance than at later ages.

Information is the only subtest which loads Factor A consistently and solely; it was the basis of the hypothetized distinction between Factors A and D. Thus, all its primary-specific variance can be attributed to verbal knowledge. Unfortunately, this is not much: .23 at age 7–6 and .13–.15 at the other ages; each of these amounts is less than the error variance at each age. For a younger subject, one might hazard an interpretation of the Information score as specific verbal knowledge *only* when it is quite deviant (at least 3 points) from the general subtest average. Even for a young subject, given the amount of error variance, this is done at considerable risk; for older children, where the primary-specific proportion is even smaller, the risk should not be taken.

As is usually the case for the WISC subtests, no subtest-specific interpretation (e.g., "this boy is unusually well informed about general facts for his ability level") should be made, given the .05–.06 of the Information variance so occupied.

As will be found to be the case in several subtests, the major utility of Information is as a measure of G and the Verbal Comprehension Factor (see below, Factor Scores) in combination with other subtests.

Comprehension

This is a moderately good measure of G compared with all the other subtests, but undistinguished when compared to the other "essentially verbal" subtests (Information, Similarities, and Vocabulary). This is undoubtedly due to the fact it has the highest error variance in this group.

In its primary-specific variance, it is of interest in that, as age increases, it goes from high to low loadings on Factor D and from zero loadings at 7–6 and 10–6 to a substantial loading at 13–6 on Factor A (Table 1). This subtest was central in the tentative interpretation of Factor D as stressing verbal judgment. Due to the fact, however, that its error variance is twice the size of its primary-specific variance, and the further fact that the latter contains mixed A and D variance at 13–6, it is not recommended that in clinical practice Comprehension scores be interpreted as measuring verbal judgment (or knowledge). The low subtest specificity also precludes any interpretation along those lines. The subtest's major utility lies in its combination with others as a measure of G and of Verbal Comprehension.

Arithmetic

Arithmetic is quite similar to Comprehension as a measure of G, better than any of the performance tests, but not as good as Vocabulary or Information. Apart from G, its primary-specific variance is small (.11–.12) and varies with age: at 7–6 and 10–6 it is Factor A variance, while at 13–6 it takes on the measurement function which it maintains until old age, Factor C, Freedom from Distractibility. Under these circumstances, and with an average of a fourth of its variance in error, interpretation of a subject's score on Arithmetic in terms of either of these factors (depending on age) is inadvisable. (Note below, however, its use in combination with Digit Span.)

For children at 10–6 and 13–6, subtest-specific variance on this test compares relatively favorably with (i.e., is a little greater than) error variance. Thus, one can draw inferences of specific Arithmetic ability or disability, but only when there is a substantial departure of the subtest score from the mean of the other (particularly verbal) subtests. Using this mean as a reference point has the effect of partialling out G, which accounts for the largest amount of the subtest's variance.

Similarities

This subtest is similar to Comprehension in that it is a moderately good measure of G (surpassed only by Vocabulary and Information), with .44 of its variance, on the average, so occupied.

At 7–6 and 10–6, its primary-specific variance comes from Factor A, while at 13–6 it is made up about equally of A and D. Again the amount is small, both absolutely and relative to the amount of error variance, and clinical interpretation along these lines is precluded. The same consideration applies to its subtest-specific variance.

The major utility of Similarities, therefore, lies in its combination with other subtests in the formation of factor scores.

Vocabulary

This subtest stands out as it has in the other analyses (Cohen: 1952a, 1952b, 1957a, 1957b) as a measure of G. On the average, G occupies .61 of its variance, and this average is depressed by the youngest group, where subtest reliabilities are consistently low. The G proportions at 10–6 and 13–6, respectively .74 and .62, correspond to correlations with G of .86 and .79. These values are quite high, and justify the frequent practice of using this subtest by itself as a basis of estimating intelligence in research and clinical screening batteries.

Because of the fact that G is so strongly verbal in character, a good measure of G such as Vocabulary has little primary-specific variance, and not much more subtest-specific variance. The source of the primary-specific variance is mixed as a function of developmental changes, going from primarily Factor D variance to primarily Factor A variance (where it remains in adulthood). Therefore, despite the relatively (and at 10–6 and 13–6 absolutely) low error variance, specific interpretations of scores are not advised. Vocabulary is essentially a G measure, used either by itself or in combination with other subtests. It can also be used, as will be seen below, in the construction of a factor score to measure Verbal Comprehension.

Digit Span

The most salient characteristic of this subtest is its poor reliability, the lowest, on the average, of the subtests. With from .40–.50 of its variance made up of measurement error, it follows that it cannot be depended upon by itself as a measure of G nor of Freedom from Distractibility (although it measures it purely). Although it has utility in combination with Arithmetic as a measure of Factor C on the WAIS in adults (Cohen: 1957a, 1957b) and in older children (see below), at ages 7–6—10–6, the Arithmetic test carries Factor A variance and Digit Span cannot be so combined.

Picture Completion

This test is one of the four outstandingly poor measures of G on the scale (the others being Object Assembly, Coding, and Mazes). It averages only .19 ($r = .43$) of its variance so occupied.

The primary-specific variance of Picture Completion is of interest from a developmental point of view. During the entire adult age range, it is the only subtest found to consistently load Factor D, which led to the consideration of this factor as a Picture Completion specific in the WAIS study (Cohen: 1957a, 1957b). In the present analysis, although Picture Completion again loads Factor D consistently, verbal subtests also load this factor, and Picture Completion has important loadings at 10–6 and 13–6 on Factor B (Table 1). Thus, the relatively large primary-specific variance at ages 10–6 and 13–6 (Table 4) cannot be attributed to any single source. Further, this subtest's error variance is the largest among the subtests, which by itself precludes specific interpretations.

Picture Arrangement

This subtest has the distinction of providing the best single measure of G among the performance scale tests, with an average proportion of G variance of .36 ($r = .60$). It is particularly good in this regard at ages 7–6 and 10–6. Although this level of G variance is exceeded by all the verbal scale subtests except Digit Span, in the testing of children with language or educational handicaps where verbal subtests are invalid, this subtest can be used, particularly in combination with Block Design, to get a nonverbal estimate of G.

High error variance and mixed sources of primary-specific variance preclude interpretation of Picture Arrangement scores along specific lines. Its utility lies in its measurement of G.

Block Design

This is a very useful subtest in its measurement both of G and Perceptual Organization. This is made possible by its low error variance, ranking with Vocabulary as the lowest on the scale when averaged over the three age groups.

Although the proportion of G variance in Block Design does not average quite as high as that of Picture Arrangement, the former is superior at age 13–6. Thus, at older ages in particular, Block Design may be a useful nonverbal measure of G, especially in combination with Picture Arrangement.

This is the first subtest thus far encountered which is found to have utility in the measurement of a primary-specific factor. Block Design loads only Factor B, and loads it quite heavily in all age groups. Thus, the proportions of variance attributed to primary-specific sources given in Table 4 (.28, .34, .26) are identifiable as almost completely due to a specific ability in speeded perceptual organization (Factor B) over and above G.

Since Factor B has the lowest correlation with G, after G is allowed for there remains a considerable amount of B primary-specific variance in Block Design. There is also about an equally substantial amount of subtest-specific variance which is likely to be largely due to spatial visualization ability.

Thus, Block Design is a good measure of G for a nonverbal test and a good measure of perceptual organization ability. This latter quality is shared with Object Assembly.

Object Assembly

This test has been consistenly found to be a poor measure of G (Cohen: 1952a, 1952b, 1957a, 1957b) and the present findings for children agree. It not only has a small proportion of G variance (.18–.23), but has a high proportion of error variance in the three groups, .37, .37, and .29. Object Assembly is therefore a very poor choice as a measure of G.

Very little of its variance is specific to it. Thus, the remaining proportion of the variance (.37, .32, .48) constitutes the largest amount of primary-specific variance among the subtests, and since its primary factor loadings are high and fall exclusively on Factor B (Table 1), all the primary-specific variance can be attributed to perceptual organization ability. Its already noted poor reliability hampers its use as a measure of Factor B by itself, but in combination with Block Design (and in younger children, Mazes), a potentially useful measure of Factor B results (see below).

Coding

Except for a minor change of content in Coding A in younger children (Wechsler, 1949, pp. 85–86), this is equivalent to the Digit Symbol test on the WAIS and gives results identical with it—it loads exclusively and consistently on Factor E, and since it is the only subtest which does so, Factor E remains uninterpreted.

Error variance is estimated (Wechsler, 1949, pp. 13–14) as being so high as to rank this test among the two or three poorest in this regard. Taken together with its relatively low G and primary-specific variance, the result is a subtest the bulk of whose reliable variance lies in its specificity (whose nature is of necessity unknown) which is in turn smaller than its estimated error variance. Taken by themselves, then, Coding scores are of limited utility.

Mazes

This is the only subtest on the WISC which has no counterpart on the W-B or WAIS. Although the proportion of its variance attributable to G in the three groups is relatively low (.14, .24, .24), it is nevertheless a useful subtest. To begin with, it is not crippled by high error variance as are most of the other performance subtests. This gives it "room" in which to measure usefully. Its primary-specific variance at the younger ages is substantial (.31 and .19) and is primarily occupied in the measurement of Perceptual Organization. In these younger years, its combination with Block Design and Object Assembly makes for a useful measure of Factor B. At age 13–6, however, only .04 of its variance is primary-specific, and this insignificant amount is not from Factor B, but from Factor C.

More by far than any other subtest (with from a third to almost a half of its variance), Mazes measures something quite specific to it. What this is cannot be determined from the present study by the very nature of subtest specificity. However, Porteus (1950), who has long been associated with this item-type, believes the subtest to measure "planning ability." A consideration of this subtest suggests that it is reasonable to identify most of Mazes' specificity with this function. This would then mean that if a child's Mazes score departed by several points from his other subtests (from Block Design and Object Assembly in particular), it is reasonable to interpret it as a consequence of planning ability, as such.

WISC IQs

Utilizing an extension of the formula for the correlation of sums given by Tryon (1958, p. 27), the correlations with G of the Verbal IQ (6 subtests), the Performance IQ (6 subtests), and the Full Scale IQ (12 subtests) were determined and squared, giving the proportion of variance attributable to G. The reliabilities of these composite IQs were found as a function of the subtest reliabilities by means of Tryon's formula for battery reliability (Tryon, 1957, Eqs. 10 and 29), and from these the proportion of error variance. These values are presented for the three age groups in Table 4, and, in addition, the residual proportions are given under the heading "Battery-Specificity," since they represent the proportion of valid non-G variance in each composite.

Both the Verbal and Full Scale IQs are excellent measures of G. About .85–.90 of their variance is so occupied, which, expressed as validity coefficients, amounts to a range of values of .92–.95. Either of these IQs represents well the G defined by the 12 subtests. Performance IQs, on the other hand, contain only .61–.68 of G variance, which, although translated into rather impressive validity coefficients of .78–.82, are relatively poor compared to the other two IQs.

The fact that the intellectual functioning represented by G is considerably dependent on verbality is indicated by the fact that the Full Scale and Verbal IQs are about equally valid as measures of G. Apparently, when the performance scale subtests are added to the verbal scale subtests to make up the Full Scale IQ, whatever G variance the former bring is counterbalanced perfectly by the non-G (specificity and error), resulting in no additional validity for G measurement beyond that supplied by the verbal scale.

Beyond G, the very little variance which is left in the Full Scale and Verbal IQs is about equally accounted for by error and primary- or subtest-specificity, the latter being obviously quite negligible for test interpretation purposes.

In the case of the Performance IQ, although the error variance is slightly larger, there remains about a quarter of the total variance specific to these tests, which is predominantly reflective of primary-specific Factor B, that part of Perceptual Organization which does not enter into G. In clinical use, the comparison of Verbal with Performance IQ for a subtest is seen to be justified, in that the difference between the two reflects the battery specificity of the Performance Scale, the effect of G having been in effect "partialled out" by subtraction. The modification of clinical practice suggested by the present results lies in the conception of the Verbal IQ as a measure of G, rather than of a "group" or primary factor of verbal ability. Another way in which this can be put is that *insofar as the WISC is concerned,* no distinction in measurement function can be made between the Verbal IQ and the Full Scale IQ for the general population of children.

Factor Scores

In the article presenting a factor-analytic rationale for the WAIS (Cohen, 1957a), factor scores for the measurement of the primary factors were advocated. The task of formulating such scores for the WAIS was simpler than proved to be the case for the WISC because of the greater consistency of subtests in their factor measurement which obtained for adults (Cohen, 1957b) than is found to be the case in children.

At the very beginning of this attempt it is noted that only Information loads Factor A exclusively and consistently, and that while Comprehension and Picture Completion both load Factor D consistently, neither does so exclusively (Table 1). Since to obtain relatively unambiguous factor scores at least two exclusively and substantially loading subtests are desirable, it is not possible to formulate separate A and D factor scores.

However, it has been pointed out above that the distinction between Factors A and D, both called Verbal Comprehension factors, is not sharp. In the light of the high correlation generally obtaining between them, and since the result was perfect consistency with the WAIS, it was decided to pool these two factors into a single Verbal Comprehension factor. When this is done, it is found that the four tests which loaded either factor both consistently and exclusively were the same as were composited into a Verbal Comprehension factor score on the WAIS, i.e., Information, Comprehension, Similarities, and Vocabulary (Cohen, 1957a, p. 456).

A factor score for Verbal Comprehension (VC) is found by averaging the WISC weighted scores for the above four subtests. For a given child, Verbal Comprehension ability is high or low relative to children his age as the VC score is above or below 10, the population mean, and high or low ipsatively (i.e., within the pattern of the child's abilities), as it departs from the child's other factor scores. The other factor scores are similarly interpreted.

Perceptual Organization factor scores (PO) must be differently made up at different ages. At ages 7–6 to 10–6, it is the average of Block Design, Object Assembly, and Mazes; at 13–6 and above, Mazes must be dropped from the composite because it ceases to load Factor B (Table 1), and the resulting PO score is identically composed as it is for the WAIS (Cohen, 1957a, p. 457).

Freedom from Distractibility scores (FD) are not available at 7–6 to 10–6, since at these ages, no subtest can be found to supplement Digit Span, which alone loads Factor C exclusively (Table 1). At age 13–6, Arithmetic can be averaged with Digit Span for an FD score, resulting in the same Factor C score as formulated for the WAIS (Cohen, 1957a, p. 457).

Finally, since all subtests are correlated with G, a G factor score is obtained as simply the average of all the subtests. This score will, like the primary factor scores above, have a population mean value of 10, and be comparable with them for purposes of ipsative measurement. If only a measure of G is required, the Full Scale IQ (or the Verbal IQ) serves this purpose. If it is not necessary to determine the conventional IQs, a short form of the WISC can be administered which yields all the recommended factor scores. The subtests necessary for this purpose are:

Ages 7–6 and 10–6: Information, Comprehension, Similarities, Vocabulary (VC); Block Design, Object Assembly, Mazes (PO); the preceding 7 subtests' mean gives the short-form G score.

Age 13–6: Information, Comprehension, Similarities, Vocabulary (VC); Block Design, Object Assembly (PO); Digit Span, Arithmetic (FD); the preceding 8 subtests' mean gives the short-form G score. Note that at this age, the factor scores are identically composed as those for adults on the WAIS (Cohen, 1957a).

The significance of these factor scores can only be determined by research. However, the logical implications of the analysis are such as to suggest that these scores, since they follow definable functional unities in children, should be of greater use than either the relatively unreliable and ambiguous single subtest scores on the one hand, or the more or less a priori Verbal and Performance IQs on the other (Cohen, 1957a, p. 456).

In order to facilitate the use of these factor scores, it is desirable that in addition to the population mean (which is 10 throughout), the variability of each factor score in the population be known. Applying to the standardization data the formula for the standard deviation of a mean (Guilford, 1950, p. 456), the values of the standard deviations of each of the above factor scores at each age group were determined and are presented in Table 5. As an approximation in clinical work, the standard deviations of the primary factor scores can be taken as $2\frac{1}{2}$, and of G factor scores (based on 12 subtests) as 2 (in contrast to subtest scores, where the *SD* is 3). These factor scores are normally distributed, so that, for example, a primary factor score of 15 can be taken as falling approximately two standard deviations above the mean and therefore in the top 3% of the population.

For ipsative (i.e., pattern) analysis, the factor scores of a subject can be compared

Table 5

Standard Deviations of WISC Factor Scores at Ages 7–6, 10–6, and 13–6

Factor Score	Age 7–6	Age 10–6	Age 13–6
VC	2.3	2.6	2.6
PO	2.4	2.4	2.7
FD	—	—	2.5
G (12 subtests)	1.9	2.0	2.0
G (short form)	1.9	2.2	2.1

among themselves much as subtest scores are. The effect of such comparison is to "partial out" the all-pervading influence of G, resulting in differences which can be attributed to what is specific to the primary factors involved.

Summary and Conclusions

The WISC standardization data for the age groups 7–6, 10–6, and 13–6 were factor-analyzed by group, using complete centroid extraction, oblique rotation to a criterion of simple structure and a positive manifold, and a second-order general factor analysis. The proportions of the variance attributable to the general factor, communality, specificity, and error were compared both among the children's groups and with adults' WAIS performance (Cohen, 1957b). The following conclusions were drawn:

1. Five correlated factors were found consistently in the three children's age groups: Verbal Comprehension I and II, Perceptual Organization, Freedom from Distractibility, and a quasi-specific factor. These are essentially the same factors which were found for adults on the WAIS.

2. A second-order general factor, G, accounted for about one-third of the total variance and about one-half of the true variance of the WISC. This factor has a very similar loading pattern to its adult counterpart, being measured best by the essentially verbal subtests.

3. Subtest specificity is relatively small, which renders invalid the clinical rationales which are dependent on distinctive measurement functions of the 12 subtests. This again duplicates the findings on the WAIS (Cohen, 1957b).

4. It was found that children exhibit a substantially smaller degree of generality of intellectual functioning than do adults. This is directly counter to the widely held belief that intelligence in children is highly general and differentiates progressively as they grow to maturity.

5. Finally, from an analysis of the sources of score variance, each subtest's measurement function was discussed in terms of G, the primary factors, specificity and measurement error. It was found, with some exceptions, that single subtest scores do not lend themselves to individual interpretation. Similar analysis of the IQs revealed the fact that both the Verbal and Full Scale IQs are excellent measures of G. Some factor scores were proposed, and the manner of their utilization was described.

Received July 25, 1958.

REFERENCES

Anastasi, Anne. The nature of psychological "traits." *Psychol. Rev.*, 1948, **55**, 127–138.

Balinsky, B. An analysis of the mental factors in various age groups from nine to sixty. *Genet. psychol. Monogr.*, 1941, **23**, 191–234.

Birren, J. E. A factorial analysis of the Wechsler-Bellevue Scale given to an elderly population. *J. consult. Psychol.*, 1952, **16**, 399–405.

Burt, C. The differentiation of intellectual ability. *Brit. J. educ. Psychol.*, 1954, **24**, 76–90.

Cattell, R. B. *Factor analysis*. New York: Harper, 1952.

Cohen, J. A factor-analytically based rationale for the Wechsler-Bellevue. *J. consult. Psychol.*, 1952, **16**, 272–277. (a)

Cohen, J. Factors underlying Wechsler-Bellevue performance of three neuropsychiatric groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 359–365. (b)

Cohen, J. A factor-analytically based rationale for the Wechsler Adult Intelligence Scale. *J. consult. Psychol.*, 1957, **21**, 451–457. (a)

Cohen, J. The factorial structure of the WAIS between early adulthood and old age. *J. consult. Psychol.*, 1957, **21**, 283–290. (b)

Cureton, E. E. A note on the use of Burt's formula for estimating factor significance. *Brit. J. statist. Psychol.*, 1955, **8**, 28.

Garrett, H. E. A developmental theory of intelligence. *Amer. Psychologist*, 1946, **1**, 372–378.

Guilford, J. P. *Fundamental statistics in psychology and education*. (2nd ed.) New York: McGraw-Hill, 1950.

Guilford, J. P. *Psychometric methods*. (2nd ed.) New York: McGraw-Hill, 1954.

GULLIKSEN, H. *Theory of mental tests.* New York: Wiley, 1950.

HOVER, G. L. An investigation of differences in intellectual factors between normal and neurotic adults. Unpublished doctoral dissertation, Univer. of Michigan, 1950.

MCNEMAR, Q. On the number of factors. *Psychometrika,* 1942, 7, 9–18. (a)

MCNEMAR, Q. *The revision of the Stanford-Binet Scale.* New York: Houghton Mifflin, 1942. (b)

PORTEUS, S. D. *The Porteus Maze Test and intelligence.* Palo Alto: Pacific Books, 1950.

RAPAPORT, D. *Diagnostic psychological testing.* Vol. I. Chicago: Year Book Publishers, 1945.

SIMKIN, J. S. An investigation of differences in intellectual factors between normal and schizophrenic adults. Unpublished doctoral dissertation, Univer. of Michigan, 1950.

THOMSON, G. H. *The factorial analysis of human ability.* (5th ed.) New York: Houghton Mifflin, 1951.

THURSTONE, L. L. *Multiple-factor analysis.* Chicago: Univer. Chicago Press, 1947.

TRYON, R. C. Reliability and behavior domain validity: Reformulation and historical critique. *Psychol. Bull.,* 1957, **54**, 229–249.

TRYON, R. C. Cumulative communality cluster analysis. *Educ. psychol. Measmt,* 1958, **18**, 3–35.

WECHSLER, D. *Wechsler Intelligence Scale for Children. Manual.* New York: Psychological Corp., 1949.

WECHSLER, D. *Manual for the Wechsler Adult Intelligence Scale.* New York: Psychological Corp., 1955.

WECHSLER, D. *The measurement and appraisal of adult intelligence.* (4th ed.) Baltimore: Williams & Wilkins, 1958.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

DIRECT MEASUREMENT OF COGNITIVE DEFICIT IN BRAIN-INJURED PATIENTS[1]

HAROLD L. WILLIAMS, ARDIE LUBIN, AND CHARLES F. GIESEKING

Walter Reed Army Institute of Research

"To measure the effects of any condition on efficiency, whether it be disorder, cerebral injury, or age, should require on logical grounds that an individual's efficiency be measured before and after the onset of the condition. In practice this nice logical requirement has not been feasible. Only very recently has any considerable proportion of the population been given the standard tests of intelligence, and no investigator, to our knowledge, has been able to obtain predisorder or preinjury test results from the patients in his sample. For this reason substitute controls have been devised" (Hunt & Cofer, 1944, p. 972).

Direct measures of premorbid intelligence are now available on all patients in the U. S. Armed Forces. The present study uses these direct measures to answer the following questions:

a. What is the amount of cognitive deficit produced by brain injury?

b. Does brain injury cause differential deficit in intellectual skills?

c. What is the best way of using the patient's premorbid and current level of intelligence to diagnose brain injury?

In the past, indirect measures of premorbid intelligence have been used to answer these questions. In dealing with individual patients, clinical psychologists have typically assumed that certain intelligence tests, such as vocabulary, were resistant to brain injury, and have used these scores to estimate premorbid intelligence. In dealing with groups, population norms have been used.

Measures

All male enlistees and inductees who come under draft quotas must take the Armed Forces Qualification Test (AFQT). The AFQT is a spiral omnibus mental test which yields a single score.

Since 1948, all enlisted personnel have taken the Army Classification Battery (ACB) upon entry into the Army. The ACB yields 10 separate scores, which, in various combinations, predict achievement at different Army schools with validities of about .60. Each test has a mean of 100 and a standard deviation of 20.

By readministering either of these tests during or after hospitalization, direct measures of cognitive deficit can be obtained (Montague, Williams, Lubin, & Gieseking, 1957). In the present study the ACB was used, rather than the AFQT, so that differential deficit could be studied.

The first five tests of the ACB were administered to every *S:* Reading and Vocabulary (RV), Arithmetical Reasoning (AR), Pattern Analysis (PA), Mechanical Aptitude (MA) and Clerical Speed (ACS). These five tests require about two-and-a-half hours. Often it was necessary to have two separate testing periods, particularly with brain-injured *S*s. Alternate forms of the ACB were used for the retests. The 10 ACB tests have previously been described in detail (Montague et al., 1957).[2]

Subjects

All *S*s were enlisted males in the U. S. Army. The brain-injured group consisted of

---

[1] Future papers will deal with the direct measurement of cognitive deficit in functional psychoses and neuroses.

[2] This study used the 1948 form of the ACB. Since June 1957, a revised battery containing 11 tests has been in use.

Table 1

Brain-Injury Diagnoses

Diagnoses	*N*	*n*
Traumatic head injuries	43	
Posttraumatic encephalopathy		15
Subdural hematoma		3
Motor seizures		3
Others		22
Cerebral vascular accident	11	
Neoplasm	3	
Miscellaneous	7	
	—	
	64	

64 patients referred by the Neurology and Neurosurgery Services at Walter Reed Army Hospital. Brain injury was defined as intrinsic damage above the tentorium cerebelli. Table 1 gives a summary of the diagnoses. Most patients had diffuse rather than focal brain damage, and the amount of brain tissue involved varied considerably. The age range was from 18 to 50, with a mean of 27 and a variance of 71.

All of the 64 brain-injured patients had recovered to a point where they could be tested, approximately two months after hospitalization. Many brain-injured patients never recovered sufficiently to be tested, and these, of course, could not be included in the study.

Several control groups were used:

a. Patient controls; 47 non-brain-injured, and nonpsychiatric patients from Walter Reed Army Hospital, and

b. Duty controls; 55 enlisted men on duty in a field hospital at Fort Meade, Maryland, and 60 enlisted men on duty in the Walter Reed Troop Command.

A detailed questionnaire was used to detect any control with a possible head injury. Such *S*s were not included in the study. The age range of the control *S*s was from 18 to 51, with a mean of 27 and a variance of 48.

Results

Table 2 presents the answer to Question *a:* What is the amount of deficit produced by brain injury? The brain-injured *S*s *decreased* about 67% of a standard deviation, whereas the controls *increased* slightly, about 15% of a standard deviation. There were no significant differences between the hospital and duty controls. For each test in the brain-injured group, Student's *t* showed that the decrease was significant at the .01 level.

The second question is a familiar one: Is there evidence of differential deficit? That is, are any of the five ACB tests relatively resistant to brain injury? In partial answer to this question, Table 2 shows that, contrary to the generalizations made by many previous investigators (e.g., Babcock, 1930; Hunt, 1943; Shipley, 1940; Wechsler, 1944), the deficit on Reading and Vocabulary is at least as great as for any of the five tests.

For a more complete answer to the question of differential deficit, we can test the null hypothesis that the average decrease in the brain-injured group on all five tests is

Table 2

Difference between ACB Premorbid and Retest Scores for Controls and Brain-Injured Patients

	Means			Variances		
Test	Brain Injured ($N=64$)	Patient Controls ($N=47$)	Duty Controls ($N=115$)	Brain Injured	Patient Controls	Duty Controls
RV	−16.5	1.4	2.8	370.6	235.3	162.0
AR	−15.1	1.5	1.4	362.0	200.5	139.3
PA	−9.2	4.3	4.0	419.0	328.6	302.0
MA	−9.8	3.2	5.2	392.4	187.5	144.4
ACS	−16.0	3.4	4.6	373.3	213.7	143.8
	—	—	—			
D	−66.6	13.8	18.0	5,141.3	1,869.8	1,386.7

Note.—D is the sum of the differences for all five ACB scores.

equal. Two approximate statistical techniques were used to test this hypothesis; both gave the same result: a significant, but *very* small tendency for PA and MA to resist the effects of brain injury.

The use of approximate statistical techniques is forced because there is no simple exact test for the differences between three or more correlated means.[3] Therefore, Box's method and Kendall's W were used.

a. *Box's* (1954) *test for differences between correlated means*. The usual two-way analysis-of-variance was calculated for the subject-by-difference score matrix, within the brain-injured group. Box's test calls for a correction to the degrees of freedom which makes an approximate allowance for the effect of intercorrelated scores as well as heterogeneous variances. The F ratio for the between-tests main effect was just significant at the .05 level, when Box's correction was applied.

One difficulty of the Box test is that it assumes multivariate normality of the five difference scores whereas, on inspection, it appears that in the brain-injured group, the scores are skewed negatively.

b. *Kendall's W test* (1948). Therefore, Kendall's rank-order test was applied to see if the tendency for PA and MA to be resistant to brain injury was relatively consistent for all 64 brain-injured *S*s. Kendall's W is a close approximation to an average Spearman rank-order correlation. (The significance test assumes that the score distribution is the same for each of the five tests, but the distribution does not have to be normal.) W equaled .06, which is, of course, very small but significant at the .01 level.

What is the best way of using the premorbid and retest scores to differentiate brain-injured and controls? For groups with multivariate normal distributions and equal covariance matrices, there are only a few possible ways to do this: (*a*) the simple difference, $T_2 - T_1 = D$, where T_2 is the sum of five retest scores and T_1 is the sum of the five premorbid scores; (*b*) the residual change score, $T_{2.1} = T_2 - a - bT_1$, which corrects for the improvement shown by controls on retest (Montague et al., 1957); or (*c*) the Fisher (1941, p. 279) linear discriminating function, the multiple regression of all 10 scores (five premorbid tests and five retests) on the dichotomous criterion, brain-injured vs. controls.

An empirical comparison of these three methods showed that the cumbersome multiple regression formula is not necessary here for optimal discrimination. The three methods are intimately related to each other in the following way: the validity of the linear discriminating function must be equal to or greater than the validity of the residual change score, which in turn must equal or exceed the validity of the difference score. The validity of the residual change score will equal that of the multiple regression if (*a*) controls and brain-injured differ on T_2, but not on T_1, (*b*) the correlation between T_1 and T_2 is substantial, (*c*) the multiple regression weights for all five premorbid tests are near zero, and (*d*) the multiple regression weights for all five retests are positive and equal. In other words, if T_1 is a suppressor variable, then the residual change score will yield optimal discrimination (Lubin, 1957).

For the computation of the Fisher linear discriminator, a special subgroup of 64 men was drawn from the sample of 162 controls so that each brain-injured patient was matched with a control within 20 points of his premorbid score, T_1. Table 3 shows that matching on T_1 tended to match all of the five premorbid scores. In general, using equal numbers and matching will maximize the difference between two groups. When the controls were given a dummy score of unity and the brain-injured a dummy score of zero, the multiple correlation of the linear discriminator with this dichotomous criterion was .66.

To obtain the residual change score, a and b were calculated, using all 162 control *S*s: $b = r_{12} s_2/s_1$, and $a = \bar{T}_2 - b\bar{T}_1$. Using the obtained values, $b = .91$ and $a = 58.05$, the validity of the residual change score was found to be .61, not significantly different from the multiple correlation.

A further simplification is possible. If r_{12} is sufficiently high and s_2 is approximately equal to s_1, then the difference score, D, will have

---

[3] Since submission of this manuscript, we have learned that Rao (1952, pp. 239–243) has an exact test appropriate to the case where the distribution is multivariate normal.

Table 3

Descriptive Statistics on Premorbid and Retest Scores

	$\bar{X}_1$		$\bar{X}_2$		Brain Injured Plus Controls ($N=128$)		
Test	Brain Injured ($N=64$)	Controls ($N=64$)	Brain Injured ($N=64$)	Controls ($N=64$)	s_1^2	s_2^2	r_{12}
RV	98.4	99.7	82.0	101.6	302.8	472.4	.56
AR	92.8	94.3	77.7	96.2	341.8	393.4	.56
PA	94.7	92.1	85.4	97.1	410.1	407.1	.42
MA	94.0	95.4	84.3	99.5	256.3	337.1	.44
ACS	84.1	85.9	68.1	90.2	334.9	428.4	.52
T	464.0	467.4	397.5	484.6	4,541.5	7,026.8	.57

optimal validity. The correlation of D with the dichotomous criterion was .60, almost equal to the validity of the residual change score, but significantly lower than the multiple correlation.

It is clear that either D or $T_{2.1}$ could be used for discrimination with no practical loss in validity. We have chosen to use the residual change score because it has a theoretical advantage as the ratio of controls to brain-injured increases.

How well does the residual change score discriminate the brain-injured from the controls? Unfortunately, there is no single answer to this question. If a residual change score of zero is used as a cutoff point, then 50% of the controls will be correctly classified, and 92% of the brain-injured will be correctly identified, the over-all percentage of correct classification for the total sample of 226 being 64. If −54 is used as a cutoff point, the over-all percentage of correct classification is 85 (the best we have done in this sample); for the controls, it is 94% correct classification; for the brain-injured, it is 62% correct classification.

One of the greatest difficulties in deciding which cutoff point to use is that, as previous writers have emphasized (e.g., Meehl & Rosen, 1955), usually the percentage of correct classification will shift if the ratio of controls to brain-injured is changed in subsequent samples. However, it is always possible to find a cutoff point such that the percentage of correct classification is invariant with respect to changes in base rates. This is the point where the percentage of correct classification in one group equals the percentage of correct classification in the other group. In our study, the cutoff point of equality of misclassification is −33, and it yields, as an invariant, 82% correct classification. This seems to be better than the percentage of correct classification reported in most studies of brain injury.

Discussion

Only two previous studies seem to have used direct measures of the premorbid intellectual level in estimating the deficit due to brain damage. Canter (1951) found that the AGCT scores of 23 multiple sclerotics decreased about three-fourths of a standard deviation. Weinstein and Teuber (1957), using the AGCT, indicate that for patients with focal gunshot wounds, there is little evidence of deficit 10 years after injury.

The Weinstein-Teuber findings could be interpreted as evidence that paper-and-pencil intelligence tests are not very sensitive to the effects of brain injury. Instead, we assume there was considerable initial deficit, but that the ambulatory patients of the Weinstein-Teuber sample regained their intellectual skills over a period of 10 years. This interpretation has the hopeful implication that brain injury need not cause permanent deterioration. Some preliminary data we have indicate that, under skilled medical care, almost half of the deficit may be regained within six months after injury. This is in general agreement with Wepman's (1951) results on the effect of training brain-injured patients.

To many clinical psychologists, the most surprising result will be that brain injury caused as much deficit in the verbal tests as in the others. It is commonly believed that verbal tests, and in particular vocabulary tests, are relatively resistant to brain injury. Why should our results differ from those reported by previous investigators?

One possible reason is that the RV test, which is of the recognition type, measures factors other than those tapped by the usual recall vocabulary. The majority of the items in RV require the *S* to read and understand a paragraph. Only a minority are standard vocabulary items. However, Montague et al. (1957) found that in a sample of Army recruits RV had a correlation of .76 with the Wechsler-Bellevue Vocabulary and Information subtests, and a correlation of .81 with the Wechsler-Bellevue Verbal scale. We have since found this same correlation of .76 between RV and W-B Vocabulary in two samples. (Army enlisted men, $N = 45$, and brain-injured patients, $N = 39$.) It is unlikely that RV contains additional common variance above that shared with Information and Vocabulary, since its correlations with the remaining eight W-B subtests are distinctly lower.

A more important question is whether our results do differ markedly from previous findings. Yates (1956) recently published an extensive review of previous studies in which he concludes ". . . vocabulary does decline in patients suffering from brain-damage" (p. 436). We are inclined to believe that those cross-sectional studies that reported less decline for vocabulary contained some bias due to the difficulty of matching controls with brain-injured on premorbid scores.

Conclusions

Traumatic brain injury results in a general deficit in intelligence test scores with little differential deficit. Reading and Vocabulary, Arithmetic Reasoning, and Clerical Speed decline slightly more than the spatial tests: Pattern Analysis and Mechanical Aptitude. These results, together with Yates' (1956) penetrating analysis, should end the myth that verbal tests such as Reading and Vocabulary are resistant to the effects of brain injury.

The residual change score is an optimal discriminator between brain-injured and controls, no differential weighting of the five ACB scores being necessary.

Summary

Qualification and classification tests can be used as premorbid measures of intelligence for all patients who have been screened by the U. S. Armed Forces. Retests on five tests of the Army Classification Battery furnished direct measures of cognitive deficit for 64 brain-injured patients. The average impairment was about two-thirds of a standard deviation. The verbal tests (Reading and Vocabulary, Arithmetical Reasoning, and Clerical Speed) were significantly more sensitive to brain injury than the spatial tests (Pattern Analysis and Mechanical Aptitude) but the differences were so small as to be of no practical significance.

A comparison of the brain-injured patients with 162 control *S*s showed that the simplest way of discriminating controls from brain-injured was to use $D = T_2 - T_1$, the sum of the retest scores minus the sum of the premorbid scores. Theoretically, the optimal discriminator is the residual change score, $T_{2.1} = T_2\text{-}a\text{-}bT_1$ (based on control group statistics), but both D and $T_{2.1}$ have a point-biserial validity of about .60 and a percentage of correct classification of about 82.

Received May 28, 1958.

REFERENCES

Babcock, H. An experiment in the measurement of mental deterioration. *Arch. Psychol.*, 1930, No. 117. Pp. 105.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems. II. Effects of inequality of variance and of correlation between errors in the two-way classification. *Ann. math. Statistics*, 1954, **25**, 484–498.

Canter, A. H. Direct and indirect measures of psychological deficit in multiple sclerosis: Parts I and II. *J. gen. Psychol.*, 1951, **44**, 3–25, 27–50.

Fisher, R. A. *Statistical methods for research workers.* New York: Stechart, 1941.

Hunt, H. F. A practical clinical test for organic brain damage. *J. appl. Psychol.*, 1943, **27**, 375–386.

Hunt, J. McV., & Cofer, C. N. Psychological deficit. In J. McV. Hunt, *Personality and the behavior disorders.* New York: Ronald, 1944. Pp. 971–1032.

KENDALL, M. G. *Rank correlation methods.* London: Charles Griffin, 1948.

LUBIN, A. Some formulas for use with suppressor variables. *Educ. psychol. Measmt,* 1957, **17**, 286–296.

MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.,* 1955, **52**, 194–216.

MONTAGUE, E. K., WILLIAMS, H. L., LUBIN, A., & GIESEKING, C. The use of Army tests for the assessment of intellectual deficit. *U. S. Armed Forces med. J.,* 1957, **8**, 883–892.

RAO, C. R. *Advanced statistical methods in biometric research.* New York: Wiley, 1952.

SHIPLEY, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.,* 1940, **9**, 371–377.

WECHSLER, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.

WEINSTEIN, S., & TEUBER, H. L. Effects of penetrating brain injury on intelligence test scores. *Science,* 1957, **125**, 1036.

WEPMAN, J. M. *Recovery from aphasia.* New York: Ronald, 1951.

YATES, A. J. The use of vocabulary in the measurement of intellectual deterioration—a review. *J. ment. Sci.,* 1956, **102**, 409–440.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

THE INHIBITION PROCESS, RORSCHACH HUMAN MOVEMENT RESPONSES, AND INTELLIGENCE:

SOME FURTHER DATA

MURRAY LEVINE, GEORGE SPIVACK
Devereux Foundation Institute for Research and Training, Devon, Pennsylvania

AND BYRON WIGHT
Veterans Administration Hospital, Coatesville, Pennsylvania

Rorschach (1942) originally observed that the human movement (*M*) response correlated with intelligence. Since then many investigators have confirmed this relationship, but the details of the relationship have received scant attention. In pursuing the relationship between the inhibition process, investigated by means of the Rorschach *M* response, and intelligence, the present authors undertook a survey of the reported relationship between *M* and various measures of intelligence. The results of this survey are presented in Table 1.[1]

It will be noted there is a striking uniformity in the magnitude of correlation between *M* and IQ. Across a variety of groups and with differing measures of intelligence, the correlations are generally low, but significant, and with the exception of two studies (Altus & Thompson, 1949; 1952) only linear relationships are reported. The median value of the *M*–IQ correlations is .26.[2] Second, we noted that only one study reported the correlation in a population of normal children (Hertz, 1934), and that no work has been reported with atypical children. Third, the two studies that dealt only with adult schizophrenics (Singer, Wilensky, & McCraven, 1956; Taulbee, 1955) both reported correlations approaching zero. Only in two other instances are such very low correlations reported. In both cases the populations represented an extremely narrow IQ range (i.e., "ranking research scientists" [Roe, 1951] and "Yale freshmen, ages 17–19" [Vernon, 1933]), where attenuation in range would militate against significance in correlation. Roe (1951) does report a positive correlation between *M* and a measure of *verbal* intelligence in this same group, but the hypothesis that *M* relates primarily to verbal intelligence is not wholly tenable. Several studies (Consalvi & Canter, 1957; Holzberg & Belmont, 1952; Singer et al., 1956; Williams & Lawrence, 1953; Wishner, 1948) report the same level of significant correlations of *M* with one or another *nonverbal* measure of intelligence. While the hypothesis of restriction in range suggests itself as an explanation for the low correlations in some instances, the lack of significant findings with schizophrenic groups by two independent investigators (Singer et al., 1956; Taulbee, 1955) is counter to the trend of reports in general and no obvious explanation suggests itself.

With these issues in mind, we undertook

---

[1] A number of other investigators report data which confirm the correlation between *M* and IQ, but their findings are not reported in the form of a correlation coefficient. Generally they are reported in the form of mean or median differences in *M* in groups which differ in intelligence. These studies uniformly confirm the relationship between *M* and intelligence, as measured in a variety of ways (Barrell, 1953; Davidson & Klopfer, 1937–38; Diers & Brown, 1951; Neff & Lidz, 1951; Palmer, 1955; Stainbrook & Siegel, 1944; Wittenborn, 1949).

[2] Because so few studies report the appropriate data, it is not possible to fully examine the effect of partialling out *R* on the *M*–IQ relationship. Such figures are included for the reader's edification in Table 1, whenever they could be obtained. In most instances, the present authors are responsible for the calculations.

Table 1

The Correlations Among M, IQ, and R in a Variety of Populations

Groups and References	N and Group	IQ Test	M–IQ	M–R	R–IQ	M–IQ·R
Normals						
Altus & Thompson (1952)[a]	100 Coll. Stu.	Ohio Psychol. Exam	.40 .19	— —	— —	— —
Altus & Thompson (1949)[b]	128 Coll. Stu.	Ohio Psychol. Exam	.41 .31	— —	— —	— —
	100 Coll. Stu.	Verbal Aptitude	.37	—	—	—
Consalvi & Canter (1957)	45 Adults—M & F, Ages 20–36	Progressive Matrices	.47	.41	.21	.41
		W-B Vocab.	.42	.41	.37	.32
Hertz (1934)	300 J. HS Stu.	Not stated	.26	—	—	—
Lotsof (1953)	30 Coll. Stu.	Ohio St. Psychol. Exam	.20	.41	.07	.19
Roe (1951)	64 "Ranking Research Scientists"	Specially devised Verbal	.22	—	.28	.32
		Spatial	.01	—	.05	.02
		Math	.14	—	.07	.13
Vernon (1933)[c]	25 "Yale Frosh," Ages 17–19	Composite score from several tests	.10	—	—	—
	48 "Harvard Students," Ages 16–23	Intelligence Test Battery	.32	—	—	—
Out patients						
Abrams (1955)	400 Veterans; psychiat. clinic pts.	W-B	.36	.44	.33	.25
Meltzoff (1956)	803 Veterans; psychiat. clinic pats.	W-B	.29	—	—	—
Wishner (1948)	20 Neurotics, M & F Ages 16–42	W-B	.21	—	.21	—
Hospitalized						
Armitage, Greenberg, Pearl, Berger, & Daston (1955)	503 Veteran psychiat. pts.	W-B	.26	—	.26	—
Beck (1932)	69 "feebleminded," M & F	S.B.	.26	—	.01	—
Holzberg & Belmont (1952)	50 M psychiat. pts.	W-B info. W-B sim. W-B pict. compl.	.33 .25 .22	— — —	— — —	— — —
Singer, Wilensky, & McCraven (1956)	100 Schizophrenic veterans	W-B Porteus mazes	−.03 .47	.41 .41	.25 .25	−.15 .41
Taulbee (1955)[d]	60 Schizophrenic veterans	W-B	Not sign.		Not sign.	
Tucker (1950)	100 Neurotic veterans	W-B	.26	—	—	—
Williams & Lawrence (1953)	97 Veterans, psychiat. pts.	WB Verbal Perf.	.30 .30	.65 .65	.51 .56	−.05 −.10

---

[a] Group Rorschach; M minus Popular M and Popular M produced r's of .40 and .19, respectively.

[b] Group Rorschach; value of .31 is based on a repeated application of the tests to the same population.

[c] $M\%$ used.

[d] r's not reported, but were obviously calculated.

the present study. First, we were interested in seeing whether roughly the same magnitude of correlation between *M* and IQ would hold in a group of abnormal adolescents, thus helping to fill in one gap in the current literature. Second, in light of the bulk of evidence for a positive *M*–IQ correlation, we wished to see whether we could confirm the negative findings of Singer et al. (1956) and Taulbee (1955) with hospitalized schizophrenics. This seemed important for the insight it might afford into the *M*–IQ relationship. Third, we were interested in attempting to confirm with different populations a previous finding that the error of reversing the mirror-image *N*, symbol for the number two in the W-B Form I digit symbol subtest, is related to general intelligence and to *M*. In earlier work (Levine, Glass, & Meltzoff, 1957) the reversal error has been described as a function of the insufficient delay or control of the tendency to write the normal *N* when presented with the task of writing its mirror image. This error has been shown to be significantly related to Rorschach *M*, a general measure of the delay function of the ego. Relating specific performances on an intelligence test to Rorschach *M* is one way of spelling out the details of the relationship between the personality variable of a delay function and general intelligence.

Subjects

The study included:

1. 155 adolescents, both sexes, residents at a school for emotionally disturbed and retarded children, with psychiatric diagnoses including chronic brain syndrome (26%), neurotic and personality disorders (51%), and a variety of schizophrenic and psychotic reactions of childhood (23%); IQ range, 41–135, normally distributed.
2. 209 adult, male, hospitalized veterans, with diagnoses of schizophrenia; IQ range 55–145, normally distributed.
3. 91 adult, male, hospitalized veterans with a psychotic diagnosis other than schizophrenia (25%), neurotic and personality disorders (39%), and chronic brain syndrome (36%); IQ range 55–135, normally distributed.
4. 132 adult, male, outpatient clinic veterans with diagnoses of schizophrenia (17%), neurotic and personality disorders (56%), and chronic brain syndrome (27%); IQ range 63–138, normally distributed.

Procedure

In all cases the values for *M*, *R*, and IQ were obtained from the clinical files of the respective institutions. The scoring of the original examiner was accepted in all instances. The Rorschach and the test of intelligence were administered individually, and a record was not accepted for the study if more than one week elapsed between the administration of the two tests. In our adult populations, and with most of the adolescents, the intelligence test used was the Wechsler-Bellevue Intelligence Scale, Form I. The remainder of the IQs were obtained from the Wechsler Intelligence Scale for Children (WISC). The IQ score used was always the full scale IQ. However, in many instances, the score represented a prorated value, rather than a value based on 10 subtests, since many clinicians administered only a short form of the test. This is not felt to be any great difficulty, since the correlation between almost any short form and the score derived from the full scale is very high.

The measure of intelligence test performance failure was the presence of one or more reversals of the mirror-image *N*, symbol for the number two in the W-B, Form I digit symbol subtest. The appropriate subtest, or the particular form of the W-B, was not administered to all *S*s, so data are not available on this point for all *S*s. However, we have no reason to expect any systematic selection of *S*s for whom the subtest was omitted.

Results

The correlations between *M* and IQ for the various populations we studied are presented in Table 2.

Our findings show low but significant correlations in each of the four populations studied. It will be noted that several of the correlations are based on populations which are heterogeneous for diagnosis. In the adult populations, there is no significant difference in *M* production between organics and other diagnoses. In the adolescent population, the median value of *M* increases as we move

Table 2

Some Further Correlations Among M, Wechsler IQ, and R

Group	N	M–IQ	M–R	R–IQ	$M\text{–IQ}\cdot R$
Hospitalized schizophrenic veterans	209	.37	.62	.23	.30
Hospitalized veterans, nonschizophrenic	91	.49	.45	.31	.41
Veterans pts, psychiat. clinic	133	.41	.60	.41	.23
Adolescents, residential school, all diagnoses, M & F	155	.20	.45	.19*	.18*

* Significant at $p = <.05 >.01$; all other r's are significant at $p < .01$.

from the diagnosis of chronic brain syndrome, through schizophrenia and then to emotional and personality disorders. However within each of these groups, when examined separately, those with more M had higher IQs than those with fewer M. We therefore do not feel that the M–IQ correlation in the group of atypical adolescents as a whole can be attributed to the heterogeneity of the population. In keeping with previous findings, there is apparently a more or less consistent relationship between M and IQ across almost any population.

It will be noted that a significant correlation is obtained in the group of adult, hospitalized schizophrenics. In fact, the figure of .37 is well above the median value of .26 of all other investigations summarized in Table 1.

The data depicting the relationship between M and reversals of the digit symbol subtest mirror-image N are presented in Table 3. The data in Table 3 are based only upon those *S*s for whom evidence of reversal or nonreversal of the mirror-image N was available. For the adult hospitalized group, those who reverse the mirror-image N have significantly fewer M than those who do not make this error. The finding continues to hold when R is controlled. In the adolescent group, reversers do not produce significantly fewer M than do nonreversers, but the direction of the slight differences in percentage is consistent with current and previous findings with adults.

Since the prior study (Levine et al., 1957) with adult outpatients found a significantly higher IQ for nonreversers, a similar comparison was made in our present populations. In the present adult population of hospitalized patients, the prior observation is confirmed. Nonreversers had a mean IQ of 101.3 ($\sigma = 16.1$), whereas reversers had a mean IQ of 93.5 ($\sigma = 17.5$). This difference is significant well below the .01 level of confidence. In the adolescent population, the mean IQ of nonreversers was 98.9 ($\sigma = 17.4$) and the mean IQ for the reversers was 91.7 ($\sigma = 20.6$). This difference approached significance at the .15 level of confidence.

Discussion

All of the obtained correlations between M and IQ are consistent with results obtained by

Table 3

Percentage of Reversers and Controls Showing Less Than Two M, Adjusted for R

	All Adult Inpatients[a]					Adolescents				
	Reversers		Controls			Reversers		Controls		
R	N	%	N	%	p	N	%	N	%	p
Not adjusted	54	83.3	151	58.3	.001	30	56.7	57	47.4	—
<19	32	90.6	89	70.8	.05	16	68.7	25	64.0	—
>20	22	72.7	62	40.3	.01	14	42.8	32	34.4	—

[a] This group is drawn from the 209 schizophrenics and 91 nonschizophrenic hospitalized patients who were administered the relevant subtest.

others on a variety of populations. The addition of the group of atypical adolescents adds one more link in the chain, and extends our ability to generalize about the *M*–IQ relationship. The consistency of correlation suggests a consistency of process in any population.

The significant correlation between *M* and IQ in our adult, schizophrenic, hospitalized group is in keeping with the general trend of past reports in other populations, and it fails to support findings reported by Singer et al. (1956) and Taulbee (1955). These authors, contrary to the general trend of past reports, found near zero correlations between *M* and IQ in groups of schizophrenics. We can see no obvious explanation for the discrepancy between our findings and theirs. In all three instances the patient populations were highly similar, and the samples sufficiently large. We can only suggest that weight of further evidence be called upon to settle the question as it concerns schizophrenics.

If the weight of future evidence supports our findings with schizophrenics, then the reliability of the specific result, and to some extent the factor analysis of Singer et al. (1956), would become questionable. Along this line we might note that Singer et al. did not draw an "intellectual factor" from their matrix of intercorrelations despite the fact that their variables included *R*, *M*, *C′*, IQ, and Porteus mazes. Singer's group did report a significant correlation between *M* and Porteus IQ. The fact that they did not draw an intellectual factor is surprising, since at least three other factor analyses (Consalvi & Canter, 1957; Lotsof, 1953; Williams & Lawrence, 1953) of Rorschach and intelligence test data have produced an intellectual factor containing *R*, *M*, and *C′* among other Rorschach variables. We wonder whether the same set of four factors found by Singer would be drawn out if an intellectual factor were produced, or if the specific relationship between *M* and IQ proved to be different in a new population of schizophrenics. It may eventually prove to be the case, as Taulbee (1955) suggests, that we cannot expect the same relationships between personality and intelligence to hold in schizophrenics. However, a great deal more evidence than that currently available would have to be marshalled to demonstrate the point. The failure to repeat the zero correlation between *M* and IQ reported by Singer does not in any way detract from Singer's other experimental results or theorizing.

Earlier work (Levine et al., 1957) presented the theory that the reversal of the mirror-image *N* was a manifestation of poor ability to inhibit or delay responses. Our finding with the adult hospitalized patients, that reversers have less *M* and lower IQs than nonreversers, duplicates prior results closely. It would seem that adequate functioning of a delay mechanism is an important element in earning a good score on the intelligence test as a whole. Both present findings and prior work indicate that the relationship between intelligence test performance and inhibition processes, as measured by *M*, holds in an adult population irrespective of the degree of illness, as indicated by hospitalized or outpatient status.

The reversal data in adolescents, however, presents a somewhat differing picture. Adolescents who tend to reverse the mirror-image *N* have a lower mean IQ (as with adults). However, the tendency to reverse does not seem to be related to Rorschach *M* (as it is related in adults), even though *M* and IQ are significantly correlated. This failure to replicate the reversal—*M* relationship in adolescents cannot be ascribed to the unreliability of the relationship, since the finding proved stable with adults. One conclusion that can tentatively be formulated is that *M* does not have exactly the same significance with regard to the delay function in adolescents as in adults. This possibility becomes more cogent in light of Litwin's (1957) failure to repeat important aspects of work on *M* and various forms of inhibition by Singer, Meltzoff, and others (Levine & Meltzoff, 1956; Meltzoff & Levine, 1954; Meltzoff & Litwin, 1956; Meltzoff, Singer, & Korchin, 1953; Singer & Herman, 1954; Singer, Meltzoff, & Goldman, 1953; Singer & Opler, 1956; Singer & Spohn, 1954) when she studied inhibition variables in younger age groups.

The present outcomes and survey of past work emphasize the relationship between a personality variable, the delay function of the ego, measured by *M*, and general intelligence.

It seems clear that the inhibition process bears a relationship to intelligence test performance. The answer to the question whether inhibition relates to the level of intellectual development in ways more subtle than simply making or not making gross errors remains for future research. Wiener (1957), also working in a personality oriented framework, demonstrated that the independently measured personality characteristic of "suspiciousness" was correlated with features of intelligence test performance and with the level of intelligence. Many have emphasized the closely entangled ties between personality and intelligence, but generally these have been treated as "nonintellective" factors of intelligence. By making use of the theoretical framework provided by the psychology of ego functions, as exemplified in the work of Rapaport, Gill, and Schafer (1945) and Fromm, Hartman, and Marschak (1954) with children, it may be possible eventually to relate concepts of intelligence and intelligent behavior within a single theoretical framework.

Summary

A median value of .26 was found in a survey of studies reporting M–IQ correlations. No studies were found dealing with atypical children, and two studies dealing with adult schizophrenics reported near zero correlations between M and IQ.

The present study reports M–IQ correlations for a group of atypical children (.20), for a group of psychiatric outpatient veterans (.41), for psychiatrically hospitalized, nonschizophrenic veterans (.49), and for hospitalized schizophrenic veterans (.37). Data on the relationship between reversals of the mirror-image N of the W-B digit symbol subtest, Rorschach M, and IQ are also reported. In the group of hospitalized adults, it was shown that reversers and nonreversers differed for both M and IQ. The findings were not fully confirmed in the group of atypical children.

It is suggested that a theoretical position relating measures of intelligence to the psychology of ego functions may eventually provide a framework to understand concepts of intelligence and personality in the same terms.

Received May 28, 1958.

REFERENCES

Abrams, E. N. Prediction of intelligence from certain Rorschach factors. *J. clin. Psychol.*, 1955, **11**, 81–83.

Altus, W. D., & Thompson, Grace M. The Rorschach as a measure of intelligence. *J. consult. Psychol.*, 1949, **13**, 341–347.

Altus, W. D., & Thompson, Grace M. Rorschach movement variables and verbal intelligence. *J. abnorm. soc. Psychol.*, 1952, **47**, 531–533.

Armitage, S. G., Greenberg, P. D., Pearl, D., Berger, D. G., & Daston, P. G. Predicting intelligence from the Rorschach. *J. consult. Psychol.*, 1955, **19**, 321–329.

Barrell, R. P. Subcategories of Rorschach human movement responses: A classification system and some experimental results. *J. consult. Psychol.*, 1953, **17**, 254–260.

Beck, S. J. The Rorschach test as applied to a feebleminded group. *Arch. Psychol.*, 1932, **21**, 136.

Consalvi, C., & Canter, A. Rorschach scores as a function of four factors. *J. consult. Psychol.*, 1957, **21**, 47–51.

Davidson, Helen H., & Klopfer, B. Rorschach statistics. Part I: Mentally retarded, normal and superior adults. *Ror. Res. Exch.*, 1937–38, **3**, 164–169.

Diers, W. C., & Brown, C. C. Rorschach organic signs and intelligence level. *J. consult. Psychol.*, 1951, **15**, 343–345.

Fromm, Erika, Hartman, Lenore D., & Marschak, Marian. A contribution to a dynamic theory of intelligence testing of children. *J. clin. exp. Psychopath.*, 1954, **15**, 73–95.

Hertz, Marguerite R. The reliability of the Rorschach inkblot test. *J. appl. Psychol.*, 1934, **18**, 461–477.

Holzberg, J. D., & Belmont, L. The relationship between factors on the Wechsler-Bellevue and Rorschach having common psychological rationale. *J. consult. Psychol.*, 1952, **16**, 23–29.

Levine, M., Glass, H., & Meltzoff, J. The inhibition process, Rorschach human movement responses, and intelligence. *J. consult. Psychol.*, 1957, **21**, 41–45.

Levine, M., & Meltzoff, J. Cognitive inhibition and Rorschach human movement responses. *J. consult. Psychol.*, 1956, **20**, 119–122.

Litwin, Dorothy. Sensory tonic theory as related to the development of inhibition and social maturity. Unpublished doctor's dissertation, New York Univer., 1957.

Lotsof, E. J. Intelligence, verbal fluency, and the Rorschach test. *J. consult. Psychol.*, 1953, **17**, 21–24.

Meltzoff, J. Personal communication, 1956.

Meltzoff, J., & Levine, M. The relationship between motor and cognitive inhibition. *J. consult. Psychol.*, 1954, **18**, 355–358.

Meltzoff, J., & Litwin, Dorothy. Affective control and Rorschach human movement responses. *J. consult. Psychol.*, 1956, **20**, 463–465.

Meltzoff, J., Singer, J. L., & Korchin, S. J. Motor inhibition and Rorschach movement responses:

A test of the sensory tonic theory. *J. Pers.*, 1953, **21**, 400–410.

Neff, W. S., & Lidz, T. Rorschach patterns of normal subjects of graded intelligence. *J. proj. Tech.*, 1951, **15**, 45–57.

Palmer, J. O. Rorschach's experience balance: The concept, general population characteristics and intellectual correlates. *J. proj. Tech.*, 1955, **19**, 138–145.

Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing*. Chicago: Year Book Publishers, 1945.

Roe, Ann. Psychological tests of research scientists. *J. consult. Psychol.*, 1951, **15**, 492–495.

Rorschach, H. *Psychodiagnostics*. New York: Grune & Stratton, 1942.

Singer, J. L. Delayed gratification and ego development: Implications for clinical and experimental research. *J. consult. Psychol.*, 1955, **19**, 259–266.

Singer, J. L., & Herman, J. Motor and fantasy correlates of Rorschach human movement responses. *J. consult. Psychol.*, 1954, **18**, 325–331.

Singer, J. L., Meltzoff, J., & Goldman, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, **16**, 359–364.

Singer, J. L., & Opler, M. K. Contrasting patterns of fantasy and motility in Irish and Italian schizophrenics. *J. abnorm. soc. Psychol.*, 1956, **53**, 42–47.

Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience type. *J. consult. Psychol.*, 1954, **18**, 1–9.

Singer, J. L., & Sugarman, D. Some TAT correlates of Rorschach human movement responses. *J. consult. Psychol.*, 1955, **19**, 117–119.

Singer, J. L., Wilensky, H., & McCraven, Vivian G. Delaying capacity, fantasy, and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.*, 1956, **20**, 375–383.

Stainbrook, E., & Siegel, P. S. A comparative group Rorschach study of southern negro and white high school and college students. *J. Psychol.*, 1944, **17**, 107–115.

Taulbee, E. S. The use of the Rorschach test in evaluating the intellectual level of functioning in schizophrenics. *J. proj. Tech.*, 1955, **9**, 163–169.

Tucker, J. E. Rorschach human and other movement responses in relation to intelligence. *J. consult. Psychol.*, 1950, **14**, 283–286.

Vernon, P. E. The Rorschach ink blot test. *Brit. J. med. Psychol.*, 1933, **13**, 89–118; 179–205; 271–295.

Wiener, G. The effect of distrust on some aspects of intelligence test behavior. *J. consult. Psychol.*, 1957, **21**, 127–130.

Williams, H. L., & Lawrence, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.*, 1953, **17**, 261–264.

Wishner, J. Rorschach intellectual indicators in neurotics. *Amer. J. Orthopsychiat.*, 1948, **18**, 265–279.

Wittenborn, J. R. Certain Rorschach response categories and mental abilities. *J. appl. Psychol.*, 1949, **33**, 330–338.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

PROGNOSIS OF SOCIAL ADJUSTMENT FOR MENTAL HOSPITAL PATIENTS UNDER PSYCHOTHERAPY

MARGARET J. RIOCH
Chestnut Lodge, Rockville, Maryland

AND ARDIE LUBIN
Walter Reed Army Institute of Research, Washington, D. C.

This study tries to answer the question: Can the Wechsler-Bellevue and the Rorschach tests predict the outcome of long-term intensive psychotherapy with very sick mental patients?

A number of empirical studies have been devoted to predicting response to psychotherapy (Windle, 1952) by means of psychological tests. In most cases, the patients were psychoneurotic, the psychotherapy was brief (rarely longer than six months), and length of stay in psychotherapy rather than outcome was the criterion. As yet there is no general agreement on the basic factors in prognosis.

In a detailed study of the Rorschach, Siegel (1948) found only three prognostic signs. Dickson (1949) found more, but his signs did not agree with those of Siegel.

Malamud and Gottlieb (1942) found that high Binet scores made it more likely that neurotics would benefit from psychotherapy, but the trend did not reach the .05 level of significance. Miles et al. (1951) found a similar but significant result using the Wechsler-Bellevue. On the other hand, Dickson (1949) found no evidence that IQ was prognostic.

These examples are characteristic of most studies. Windle (1952) reports some 30 studies, covering a wide variety of therapies, in which cognitive tests were used as predictors. In one third of them, the intelligent patients were more likely to improve; in another third the opposite was true; the remaining third showed no significant relation. It seems then that no consistent relation has been demonstrated between intellectual level and the results of psychiatric therapy.

The present study was conducted at Chestnut Lodge, Inc., a private hospital of about 100 beds, devoted to psychotherapy. It is important to describe the milieu since it is unique. No shock or psychosurgical treatments are used at this hospital. Well over half the patients are schizophrenics, many of whom have been hospitalized for years before coming to Chestnut Lodge. Socioeconomic status and intellectual level tend to be high. Every patient is seen regularly by an analytically trained psychiatrist, usually four hours a week. The basic philosophy of the hospital includes the assumptions that patients may require years of hospital treatment; that they probably will continue in therapy after leaving the hospital, often for several years; and that all patients, no matter how deeply psychotic, can recover.

The high economic status of the patients, the long intensive psychoanalytic therapy, and the general hospital milieu make this sample a homogeneous group of mental patients of a type never studied before from a prognostic point of view.

Method

Subjects

The *S*s of this study were all the patients who met the following conditions:

a. A Rorschach was given at Chestnut Lodge, or a Rorschach protocol taken within the six months prior to admission was available.

b. The patient was in treatment at Chestnut Lodge at least two years *following* administration of the test. This is important because the main purpose of this study is to answer the question whether tests have prognostic significance, and the prognosis must be made, of course, from the time of testing forward.

c. The patient did not have any known brain damage.

d. One of the patient's treatment years postdated January 1, 1950, and all the requirements listed above were fulfilled by October 1, 1955. No information following that date was used.

Ninety-three patients (40 males and 53 females) fulfilled all the conditions. The median age of the males was 26 and of the females, 28, with a range from 16 to 53. Table 1 gives their distribution according to diagnostic category.

Table 1

Psychiatric Classification of Patients Used in This Study

Diagnoses	*N*
Schizophrenia, paranoid type	43
Schizophrenia, catatonic type	10
Schizophrenia, hebephrenic type	6
Schizophrenia, simple type	1
Schizophrenia, other types	10
Manic depressive psychosis	2
Schizoid personality	1
Psychopathic personality	2
Personality disorder	2
Psychoneurosis	16
	93

Criterion

The really difficult part of any prognostic study is to define precisely (*a*) what one is trying to predict, and (*b*) for what time one is trying to predict it.

a. No attempt is made in this paper to predict personality change per se, or any of the psychological conditions usually conceived of as desirable end-states in psychotherapy. Only the patient's social adjustment, as shown by where he lived, was considered. The most important distinction is whether the patient was in or out of a mental hospital. If he was in a hospital, was he on a disturbed ward or a convalescent ward? If he was out of a hospital, was he an outpatient, a private patient, or was no treatment needed? Changes from one such status to another constituted our criterion of improvement. Graphs were made for each patient showing his status for each day using the following status scores:

5: Disturbed ward
4: Semi-disturbed ward
3: Convalescent ward
1: Outpatient
½: Private patient
0: No treatment needed

b. Whether a patient seems to do well or poorly depends upon the period of time over which he is observed. The term "final outcome" is either meaningless or misleading in psychiatry. Does it refer to the patient's state of mental health at the time of his death? Any other point in his life has to be chosen arbitrarily and is often followed as well as preceded by considerable change for the better or for the worse. A study of the results of therapy should, therefore, include a description of the patient's *course* for a stated period of time and whether it is upward or downward, or upward *and* downward. Clearly, no simple function of the status scores can represent adequately a patient's course. Therefore, 11 categories were set up which formed the social adjustment criterion.

Category 11 denotes the best outcome; Category 1 denotes the worst. (The numbers in brackets refer to the number of patients in the category.)

11. All scores between 0 and 1, i.e., patient was never hospitalized [2].
10. Scores begin at 3 or more and are down to 1 or less by the end of 1 year, i.e., the patient starts in the hospital, improves consistently and is out of the hospital by the end of 1 year [9].
9. Scores begin at 3 or more, are down to 1 or less by the end of 1 year, as in Category 10, but the decrease is inconsistent, up and down [0].
8. Scores begin at 3 or more, are down to 1 or less between the end of the first and the end of the third years, i.e., the patient starts in the hospital, improves consistently, but more slowly and is out of the hospital by the end of the third year [11].
7. Scores begin at 3 or more, are down to 1 or less between the end of the first and the end of the third year, as in Category 8, but the decrease is inconsistent, up and down [6].
6. Scores begin at 3 or more and are down to 1 or less after the end of 3 years, i.e., the patient starts in the hospital, improves consistently but very

slowly and is not out of the hospital until after the end of 3 years [3].

5. Scores begin at 3 or more and are down to 1 or less after the end of three years, as in Category 6 but the decrease is inconsistent, up and down [10].
4. Scores begin at 3 or more; there is no *steady* decrease, but the monthly averages are 1.6 or less for at least 3 months at the end, i.e., patient's course is inconsistent, but there is a trend toward improvement at the end of the period [7].
3. Scores begin at 3 or more. There is no steady decrease, i.e., patient's course is inconsistent with no trend toward improvement at the end of the period [22].
2. Scores do not go below 3, but rarely go higher, i.e., the patient remains consistently hospitalized but chiefly on a convalescent ward [7].
1. Scores do not go below 3 and are for the most part higher than 3, i.e., the patient remains consistently hospitalized, chiefly on a disturbed ward [16].

Three psychologists made independent classifications with almost perfect agreement. Ninety-five per cent of all patients were placed in the same category by all three raters.[1]

Reliable information on the patient was available for periods of two to nine years. This is not necessarily all time in treatment, but simply time over which there is reliable information. The patients were divided into four groups according to the length of time during which information was available about them after the testing date. (The distribution is given in Table 3.)

Prediction Measures

Wechsler-Bellevue Adult Scale of Intelligence. A Wechsler-Bellevue, Form I or II, had been administered to 90 of the 93 *Ss* close to the time of the administration of the Rorschach. Only the Full Scale IQ, calculated on the basis of at least 10 subtests, was used in this study.

Rorschach. Preliminary studies on the correlation of particular Rorschach factors with the patients' clinical courses were unpromising. The senior author, therefore, developed the following broad descriptive statements about good signs and poor signs in the test:

Good Signs

1. Patient goes at the test competently. This may be shown in a variety of ways, e.g.:
   - *a*. He gives a large number of R which are *not simply* empty perseverations or repetitions of his preoccupations, but are of good quality.
   - *b*. He does not hesitate a great deal, but is fairly speedy in his productions.
   - *c*. He speaks as if he knows what he is talking about.
   - *d*. He shows interest, alertness, and/or enjoys the task, and/or shows some humor (not of the unconscious variety).

2. The patient's test does not show *clear* evidence of manifest psychosis. It might be diagnosed neurosis (hysteric or obsessional), schizoid, or latent schizophrenia. Mild psychosis is in between good and bad.
   - *a*. As a corollary of this there is a good number of P.

3. The content is lively, imaginative, varied, original. Dysphoric, morbid content is not the best but is much better than dull content.

4. There is good movement, color, shading, but particularly good M.

5. The patient seems to be intelligent *and* able to function intelligently at present. This may seem to be the same as the points above, but any additional evidence of it is a good sign.

Poor Signs

1. Patient shows signs of incompetence in fulfilling the task. This may be shown in a variety of ways, e.g.:
   - *a*. He gives a meager quantity of R.
   - *b*. He rejects one or more cards.
   - *c*. He takes an inordinately long time.
   - *d*. He expresses or shows extreme fatigue, disinterest, restlessness or the like.
   - *e*. He is very uncertain of what he perceives. Try to distinguish this from the neurotic's uncertainty about whether he is pleasing the examiner.

2. The patient shows clear signs of overt psychosis. This is especially bad if the signs are so clear that a person who is not a Rorschach expert, reading the protocol, could identify it as a psychotic production simply from the verbalizations. It is also bad if the Rorschach worker finds the diagnosis of overt psychosis easy to make on a casual skimming of the protocol. The easier the diagnosis, the worse the prognosis.
   - *a*. As a corollary of this there is low P.

3. The content of the record is dull, empty, repetitive.

4. There is little movement, color, or shading.

5. The patient seems to be *either* of poor intellectual endowment *or* shows at present gross intellectual impairment.

---

[1] The authors are indebted to Harold L. Williams, Chief of the Department of Clinical and Social Psychology, Walter Reed Army Institute of Research, for help in this classification.

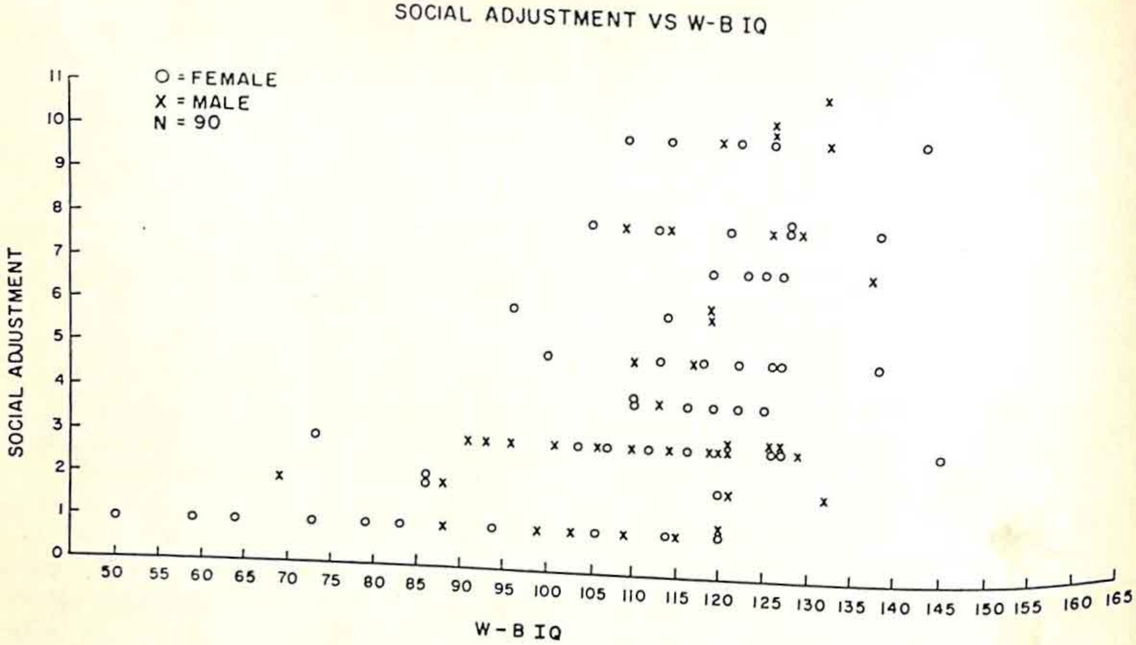


Fig. 1.

The statements were given to two experienced Rorschach workers,[2] together with the Rorschach protocols of the 19 patients in Group D (7–9 years of follow-up). They had no information about the patients other than age and sex. When it was found that their independent global ratings had substantial correlations with each other and with the social adjustment criterion, all 93 Rorschach protocols were turned over to a competent but relatively inexperienced clinical psychologist.[3] Eight protocols were used for practice and training, leaving a total of 85 protocols for scoring according to the "good and poor signs." A high score meant a predominance of good signs and vice versa.

Results

Figure 1 shows the relation of social adjustment to the W-B intelligence score. Clearly, it is not a bivariate normal distribution, but is the triangular or fan shape frequently met in psychological studies. Although not given here, the plot of social adjustment against the Rorschach rankings showed the same heteroscedasticity. Figure 1 shows clearly that accuracy of prediction from Wechsler IQ increases as the IQ decreases. For *S*s with an earned IQ higher than 110, the predictions of social adjustment are no more accurate than would be obtained by chance.

For the males, the correlation of W-B IQ with social adjustment is .50, and for the females is .55. For the total group the correlation is .52. Because of heteroscedasticity, the usual tests of significance were not made. Various rank-order and chi-square tests showed that the relation was significant.

Table 2

Intercorrelations of Social Adjustment, Rorschach Ratings, and Intelligence for Patients with 7–9 Years of Treatment Plus Follow-up Time

		M	N	E	W
Ratings by Rorschach workers					
Experienced	M				
Experienced	N	.85			
Inexperienced	E	.73	.84		
Wechsler-Bellevue IQ	W	.79	.67	.52	
Social adjustment criterion	C	.75	.76	.75	.69

---

[2] The authors are indebted to Margaret Thaler Singer and Nathene Loveland, Department of Clinical and Social Psychology, Walter Reed Army Institute of Research.

[3] The authors are indebted to Eugene Stammeyer, Department of Psychology, St. Elizabeth's Hospital.

Table 3

Intercorrelations of Social Adjustment, Intelligence, and Rorschach Rating by Follow-up Group

Group	Treatment Plus Follow-up Time	N	r_{ce}	r_{cw}	r_{ew}	$R_{c.ew}$	Wechsler IQ Mean	Wechsler IQ Variance
A	2 Years	17	.51	.26	.74	.54	112	408
B	3–4 Years	18	−.05	−.16	−.04	.17	119	67
C	5–6 Years	31	.44	.65	.64	.65	105	442
D	7–9 Years	19	.75	.69	.52	.83	113	246
		85						

Note.—c is the social adjustment criterion, e is the Rorschach rating, and w is the Wechsler IQ; $R_{c.ew}$ is the multiple correlation of the criterion with the Rorschach and Wechsler.

A premorbid W-B IQ was estimated on the basis of the W-B protocol. A deficit score was computed by subtracting the earned IQ from the estimated IQ. This index of deterioration had a correlation of about .50 with social adjustment, the same as the earned IQ.

Table 2 gives the results of applying the Rorschach global rating, described previously, to Group D (7–9 years of treatment plus follow-up time).

It seems that the two experienced Rorschach workers produce higher correlations, not only with each other but also with the Wechsler IQ, than did the inexperienced clinician.

Table 3 tries to answer several questions. Does the predictability of the criterion change as follow-up time increases? Will joint use of the Rorschach rating and Wechsler IQ improve prediction over use as individual scores? There seems to be no systematic change in predictability as follow-up time increases. The zero-order correlations make it clear that the Rorschach and the Wechsler predict with almost equal accuracy. The multiple correlations indicate that a linear combination does not improve the prognosis significantly.

In Table 3, Group B (3–4 years of treatment plus follow-up time) obviously differs from the others in that all correlations are near zero. But essentially this situation does not differ from that in other groups, for in Group B there is just one S with an IQ lower than 110, and Fig. 1 shows that it is only for such Ss that accurate prediction can be made. The lowered variance on the Wechsler thus seems to account for the near zero correlations.

Discussion

The most striking and practically useful finding in this study is that the tests, as used here, do not predict social adjustment for patients who do well on the psychological tests, but do predict with virtual certainty for those who do poorly. Whether this poor performance is a function of inferior native endowment or the result of so-called psychotic deterioration or a combination of the two, is a question which cannot be answered from this study.

In practice, the clinician utilizing long-term intensive psychotherapy with very sick mental patients will be confronted with two types: (*a*) those whose intellectual functioning is relatively intact; who approach a task involving creative imagination with alert competence; and who, if they are psychotic, still maintain "clear" areas, and (*b*) those who show obvious intellectual impairment and obvious all-pervasive psychosis, whose productions are meager, dull, and repetitive. The patients in the first group may or may not recover. The factors which lead to recovery could not be identified in this study.

The patients in the second group did not recover. The possibility cannot be excluded that continuing treatment over a still longer period of years may produce better results, but the conclusion remains that such results occur only when the effort is enormous and the case is exceptional. All of these patients were schizophrenics; their characteristics are like those which have been described in the literature as process schizophrenia. The clinician must ask himself whether psychotherapy, as it is now conceived, is really the method of choice with these patients since the results over such a long period of time are so meager.

Since no consistent relation has been demonstrated between intellectual level and outcome when all types of therapy are considered together (Windle, 1952), the high prognostic accuracy in this study for patients with low intelligence is probably linked to the kind of patient, the kind of milieu, and the kind of psychotherapy administered.

Summary

Ninety-three patients undergoing intensive psychotherapy for at least two years were given psychological tests and followed for additional periods up to seven years. Either the Wechsler-Bellevue or the Rorschach could be used to predict the patient's course with a validity of about .50; combining the two did not improve prediction. A prognosis of nonimprovement could be made with virtual certainty on the basis of poor test performance, but good test performance was not prognostic.

Received May 28, 1958.

REFERENCES

Dickson, L. P. The Roschach Test as an indicator of prognosis in psychoneurosis. Unpublished master's thesis, Catholic Univer., 1949.

Malamud, W., & Gottlieb, J. S. Course and prognosis in the psychoneuroses. *J. Mt. Sinai Hosp., N. Y.*, 1942, **9**, 630–639.

Miles, H. H. W., Barrabee, E. L., & Finesinger, J. E. Evaluation of psychotherapy with a follow-up study of 62 cases of anxiety neurosis. *Psychosom. Med.*, 1951, **13**, 83–105.

Siegel, M. G. The Rorschach Test in diagnosis and prognosis. *Family*, 1946, **27**, 102–106.

Windle, C. Psychological tests in psychopathological prognosis. *Psychol. Bull.*, 1952, **49**, 351–482.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

MMPI *L–F–K* RELATIONSHIPS WITH CRITERIA OF BEHAVIORAL DISTURBANCE AND SOCIAL ADJUSTMENT IN A SCHIZOPHRENIC POPULATION

LEONARD R. GROSS[1]
Veterans Administration Hospital, Roanoke, Virginia

In recent years there has been a change in the clinical usage of the Minnesota Multiphasic Personality Inventory (MMPI) validity scales. Since the advent of pattern interpretation, the original use of the *L–F–K* scales as indicators of test-taking attitudes has broadened to deal with such personality variables as self-concept, reality-testing, adequacy of social behavior, degree of behavioral disturbance, and general adjustment mechanisms.

Examination of data from various studies suggests that *F* is positively related to behavioral disorganization (Gough, 1946; Schmidt, 1945). Schiele and Brozeck (1948) report that *F* tends to increase with severity of disturbance following semistarvation. Gough (1950) indicates that *S*s attempting to present themselves as disturbed have a high *F* and low *K*, while *S*s told to make a "good impression" have a high *K* and *L*. Dragow and Barnette (1957) report that *S*s motivated to indicate their capabilities as employees worthy of advancement had a high *K* and low *F*.

The purpose of this study is to investigate the relationships among the validity scales with regard to some personality variables.

Since this study is exploratory, only two basic hypotheses were made, utilizing current clinical practice as well as the original logic of the scales:

1. *S*s relatively nonbehaviorally disturbed have $L > F$, $K > F$ (a "V" shape).
2. *S*s severely behaviorally disturbed have $F > L$, $F > K$ (a "caret" shape).

Method

A group of 100 newly admitted psychiatric patients, 50 males and 50 females with a primary diagnosis of schizophrenia, who were given the MMPI during the first month after admission, were selected as *S*s. Patients over 60 years of age or with a history of organic brain damage were excluded.

The mean age of the male *S*s was 33 years, with a range from 22 to 55. The mean age of the female *S*s was 34 years, with a range from 21 to 51. The mean education of the male *S*s was 11.6 years and of the females 11.8. For the males, the subclassifications of schizophrenic reactions were: 1 catatonic, 20 paranoid, 1 affective, 28 undifferentiated. For the females the subclassifications of schizophrenic reactions were: 2 catatonic, 1 hebephrenic, 18 paranoid, 4 affective, 25 undifferentiated. There were no significant differences in any of the above between sex groups.

The *S*s were rated by the investigator on scales of *behavioral disturbance* (BD) and *social adjustment* (SA).[2] The scales suffer from the fact that they had to be formulated to deal with existing clinical record data. BD was determined by rating the *S* as -1, 0, or $+1$ on 9 categories with a score of -1 being in the disturbed direction. The general categories were ward adjustment, obvious behavioral manifestations of psychosis, and therapy

---

[1] The author wishes to acknowledge his indebtedness to William J. Eichman for his aid in planning the study and to Burke M. Smith, Earl G. Guyer, and Roy A. Eck for their critical reading of the manuscript. Thanks also to the Medical Illustration Service for making the graphs.

[2] A table giving the complete scales has been deposited with the American Documentation Institute. Order Document No. 5888, remitting $1.25 for microfilm or $1.25 for photocopies.

Table 1

Analysis of Variance of *L–F–K* Relationships with Mild and Severe Behavioral Disturbances and with Male and Female Schizophrenics

Source	*df*	Mean Square
Between Subjects	99	102.59
Behavioral disturbances	1	216.75
Sex	1	234.47
BD × S	1	99.38
Error (b)	96	100.51
Within Subjects	200	133.36
Validity scales	2	617.10*
V × BD	2	1,418.57**
V × S	2	27.43
V × BD × S	2	214.33
Error (w)	192	115.18
Total	299	

* *F* significant at .01.
** *F* significant at .001.

Table 2

Analysis of Variance of *L–F–K* Relationships with Good and Poor Social Adjustment and with Male and Female Schizophrenics

Source	*df*	Mean Square
Between Subjects	99	102.59
Social Adjustment	1	298.01
Sex	1	234.47
SA × S	1	42.17
Error (b)	96	100.24
Within Subjects	200	133.36
Validity Scales	2	617.10*
V × SA	2	50.08
V × S	2	27.43
V × SA × S	2	971.20**
Error (w)	192	121.85
Total	299	

* *F* significant at .01.
** *F* significant at .001.

assignment adjustment for the 30 days following the administration of the MMPI. The SA criterion was found by similar ratings on 9 categories dealing with educational, marital, and financial adequacy, family background, and past history of emotional disturbances. The pluses and minuses were summed and a distribution of the final scores on BD and on SA were formed. Each distribution was then divided at the mean, the lower scores falling into the "severe BD" category and the "poor SA" category respectively. The higher scores were placed into the "mild BD" category and the "good SA" category, respectively.

The scores on the BD and SA scales were compared with the scores on the validity scales of the MMPI. The statistical designs used were Type III three-dimensional and Type I two-dimensional analysis of variance designs as recommended by Lindquist (1956).

Results

A three-dimensional analysis of variance was done which included the variables of (*a*) mild and severe BD, (*b*) Sex, and (*c*) the Validity scales. The results of this analysis are presented in Table 1. In this design, the variables of BD and Sex, taken alone, have no psychological meaning, since the scores dealt with are summations of the *T* scores across *L–K–F* scales. Likewise, there is little interest in the effect of the Validity scales taken alone. In this case, the Validity scale's effect is significant, and further analysis indicates that for the total sample $F > L$

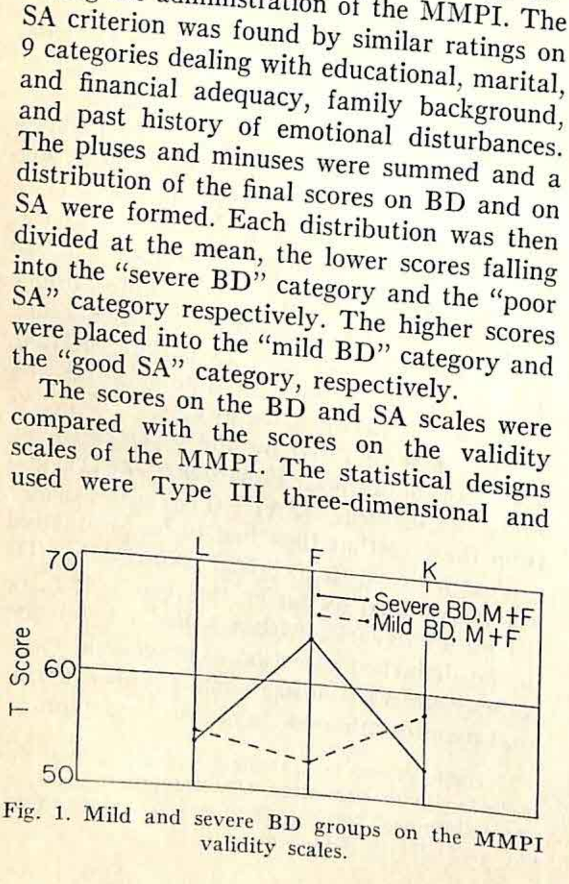


Fig. 1. Mild and severe BD groups on the MMPI validity scales.

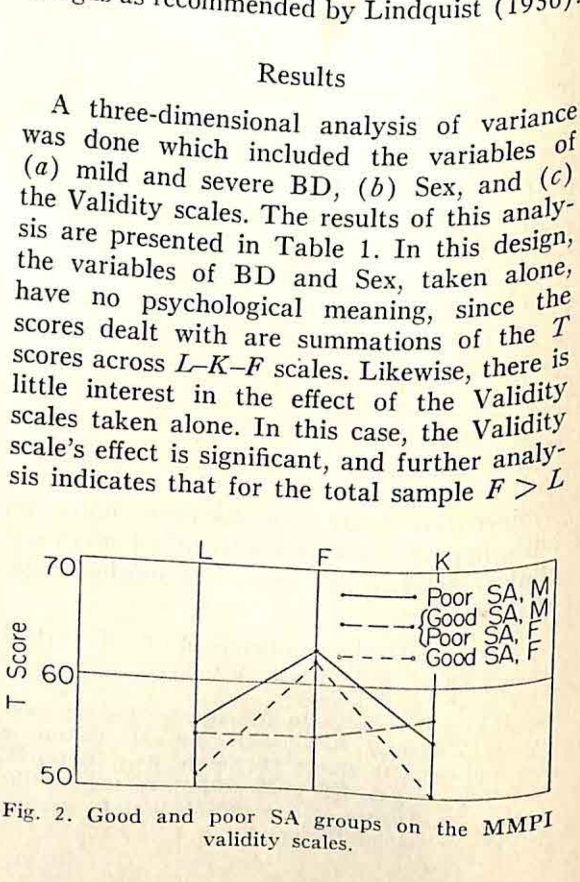


Fig. 2. Good and poor SA groups on the MMPI validity scales.

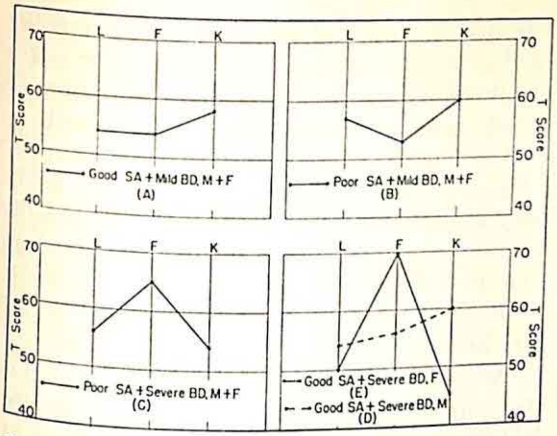


Fig. 3. SA and BD profile combinations on the MMPI validity scales.

or *K*. This is to be expected in a schizophrenic sample. Prime interest lies in the interactions of the Validity scales with Sex and with BD or with both. In this analysis the triple interaction is not significant, and there are no significant interactions with the sex of the *S*s. The difference between mild and severe BD for the Validity scales was significant beyond the .001 level. Figure 1 represents this significant interaction. Hypothesis I, that *S*s with mild BD have $L > F$, $K > F$, is confirmed. Hypothesis II, that *S*s with severe BD have $F > L$, $F > K$, is also confirmed.

An identical analysis of variance design for SA and the three validity scales also yields two significant *F* ratios, one for the validity scales beyond .01 as in BD and the other a triple interaction among the validity scales, SA and sex, which was significant beyond the .001 level. Table 2 presents the summary table and Fig. 2 represents the significant interaction. As can be seen, male *S*s with good SA have a similar configuration to female *S*s with poor SA.

Table 3

Profile Combinations (see Fig. 3) Significant Beyond the .05 Level

	A	B	C	D	E
A[a]		—	<.025	—	<.001
B[a]			<.001	—	<.001
C[a]				<.05	<.05
D[a]					<.001
E[a]					

[a] A = Good SA and mild BD, M + F; B = Poor SA and mild BD, M + F; C = Poor SA and severe BD, M + F; D = Good SA and severe BD, M; E = Good SA and severe BD, F.

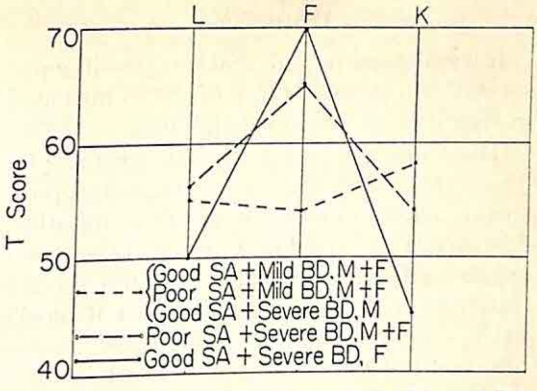


Fig. 4. Summary profiles of significant SA and BD combinations on the MMPI validity scales.

Because of the discrepancy between males and females on SA, two-dimensional analyses were run between SA and BD for the two sexes in order to ascertain whether or not males and females could be combined for any BD, SA combination. The four pairs of profiles compared for sex differences were good SA and mild BD, good SA and severe BD, poor SA and mild BD, poor SA and severe BD. This was possible to do since product-moment correlations between BD and SA for the males and females yielded an *r* of .14 and .07, respectively, indicating that the two scales were measuring different variables. The only difference found between the two sexes was on good SA and severe BD, significant beyond the .001 level. Therefore the eight profiles could be reduced to five, three in which males and females had similar profiles, and two in which the males and females had dissimilar configurations. The five profiles are diagramed in Fig. 3.

Two-dimensional analyses were then made between all combinations of the five profiles in Fig. 3 to determine if any of them differed significantly from any other. Table 3 indicates the results of the 10 analyses. On the basis of these analyses, Profiles A, B, and D in Fig. 3 were combined, leaving three distinct profiles as in Fig. 4. These were Profiles C, E, and a category in which A, B, and D belong.

Discussion

It can be suspected that the small *n* for several of the profiles in Fig. 3 contributed to their lack of statistical stability.

The three significant profiles generally indicate that the *F* scale is in large part a positive function of behavioral disorganization. Thus a high *F* and low *L* may indicate overt expressions of pathological feelings as the relatively high *F* of delinquents in Rempel's (1958) study suggests, rather than dissembling or test-taking attitude as such.

The poor SA and severe BD males and females have a high *F* with a lower *L* and *K*, although both the latter two are above the mean. A clinical interpretation supported by the data would be that an individual with such a configuration is attempting to present himself as socially acceptable, conforming and adequate (*L*) and is denying social difficulties as well as attempting to utilize repressive devices (*K*). Because of severe and chronic pathology he cannot separate normal from bizarre responses, hence the high *F*. In short, this would be a chronically disturbed individual who is attempting both consciously and unconsciously to be defensive but is incapable of doing so.

In contrast with this profile is the good SA and severe BD female *S*s who have an *L* and *K* below the mean as well as a high *F*. This group can be considered acutely disturbed as they have had good social adjustment in the past. They are blatantly admitting to pathology and revealing a poor self-concept, lability, and dysphoria. Thus it would be expected that female patients who have made a somewhat adequate social adjustment in the past are less likely to be defensive and guarded regarding their current disturbance on either a conscious or unconscious level.

The other profile, a combination of A, B, and D, is composed of three groups that score in the more healthy direction on one or both of the criterion scales. This group seems to have some assets that do not exist in the poor SA and severe BD group. It can be seen from Fig. 3, Profile A, that the good SA and mild BD group most nearly represents a straight line. They have been able to maintain some balance between self-concept, a conscious desire to perform adequately, and overt behavior. Their attempt to present themselves as adequate, nondisturbed, and conforming to social expectations is somewhat successful.

It would appear from inspection of the poor SA groups that they tend to have a relatively high *L* and *K*. Their reliance on denial of illness, inadequate repressive mechanisms, as well as their rigidity, would indicate poor interpersonal relations contributing to poor social behavior.

One of the most striking differences statistically is between the good SA and severe BD males and females. The females, as pointed out before, are admitting freely to pathology and acting out their symptomatology. The males on the other hand present a picture of both conscious (*L*) and unconscious (*K*) denial and constriction of behavior. Perhaps males that become acutely disturbed tend towards withdrawal and intratensive behavior, while females tend toward a more overtly emotional behavior pattern.

It seems worth emphasizing that even within a relatively homogeneous diagnostic group the three "validity" scales differentiated *S*s on crude a priori social-behavioral criteria and can be considered personality measures.

Because this study is exploratory, it perhaps raises more questions than it answers. While it is hoped that these three MMPI scales will be considered as representing important variables and be used as an index of adjustment mechanisms and overt behavior as well as "validity" scales per se, this study has only hinted at two possible dimensions. The sample, as noted before, was not adequate to explore fully what differences, if any, exist among the three profiles in Fig. 3, which were grouped together. The BD and SA scales were operationally defined and based on existing clinical records. Hence, they were limited by that data which was common to all patients.

Summary

At the Roanoke VA Hospital, 50 male and 50 female schizophrenic patients were rated on a priori scales of behavioral disturbance (BD) and social adjustment (SA) from their clinical records. They were divided into poor

and good SA, mild and severe BD, and then compared on the *L–F–K* scales of the MMPI. It was found that *S*s with severe BD had a "caret"-shaped *L–F–K* profile, while *S*s with little BD had a "V"-shaped profile. When BD was combined with SA, three distinct profiles emerged, indicating that the "validity" scales seem to measure personality characteristics.

Received June 2, 1958.

REFERENCES

Drasgow, J., & Barnette, W. L. $F-K$ in a motivated group. *J. consult. Psychol.*, 1957, **21**, 399–401.

Gough, H. G. Diagnostic patterns on the MMPI. *J. clin. Psychol.*, 1946, **2**, 23–37.

Gough, H. G. The *F* minus *K* dissimulation index for the MMPI. *J. consult. Psychol.*, 1950, **14**, 408–413.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1956.

Rempel, P. P. The use of multivariate statistical analysis of MMPI scores in the classification of delinquent and nondelinquent high school boys. *J. consult. Psychol.*, 1958, **22**, 17–23.

Schiele, B. C., & Brozek, J. "Experimental neurosis" resulting from semi-starvation in man. *Psychsom. Med.*, 1948, **10**, 31–50.

Schmidt, H. O. Test profiles as a diagnostic aid: The Minnesota Multiphasic Inventory. *J. appl. Psychol.*, 1945, **29**, 115–131.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

THE INTERACTION AMONG ANXIETY, STRESS INSTRUCTIONS, AND DIFFICULTY[1]

GERALD WIENER
Rosewood State Training School, Owings Mills, Maryland

This experiment is concerned with the effects of anxiety, stress instructions, and difficulty of stimulus material. An attempt is made to extend the theory and empirical findings regarding these variables to performance on projective devices.

Experiments concerning the effect of stressful situations and subjects' anxiety level on learning and performance have been reviewed by Lazarus, Deese, and Osler (1952), Child (1954), and Taylor (1956). It has been generally found that when the level of difficulty within a task is varied, *S*s' anxiety level does not hinder and sometimes aides functioning on simpler aspects of the task used. It has also been found that *S*s' anxiety and stress instructions interact so that high test anxious (HA) *S*s are more impaired in stress situations than low test anxious (LA) *S*s.

The theory used here to understand these findings has been stated by Mandler and Sarason (1952), Sarason, Mandler, and Craighill (1952), and Child (1954). Briefly the theory may be outlined as follows:

1. HA *S*s tend to react to test situations with achievement-related, anxiety-reducing responses. Anxiety-reducing responses are largely task inappropriate. They may be self-relevant, aggressive, or in other ways competing and incompatible with the tasks. To this extent, they impede performance.

2. LA *S*s tend to react to test situations primarily with task appropriate responses. LA *S*s do not tend to make inappropriate anxiety-reducing responses in a test situation.

3a. When stress is introduced via ego-involving instructions, expectations of failure and punishment are aroused along with achievement motivation. As a drive-producing cue, stress instructions serve to elicit responses which a given *S* has previously learned to make to stressful test situations. HA *S*s respond with anxiety, related to the achievement motive in conflict with expected failure and punishment. Thus, HA *S*s in a stress situation tend to have greater anxiety and strengthened anxiety-reducing responses.

3b. LA *S*s have their prepotent mode of responding to a test situation strengthened by stress. For such *S*s, stress does not serve as a cue for the introduction of anxiety and anxiety-reducing responses, but to increase achievement motivation and strengthen task appropriate responses.

4. Increasing the difficulty of a problem produces a conflict situation. Difficulty, as here used, has been defined by Farber and Spence (1953) as a function of the number of incompatible and competing response tendencies *S* must choose among in performing on a given unit of a task.

4a. HA *S*s react to difficult, conflict-producing test situations with anxiety-reducing reactions.

4b. Increasing task complexity for LA *S*s does not induce anxiety and anxiety-reducing responses.

Assumptions 4, 4a, and 4b lead to the prediction that *S*s' anxiety, stress instructions, and task complexity interact. HA *S*s compared to LA *S*s should therefore show greater impairment as test materials become more difficult, and this effect should be increased by stress instructions. This prediction was tested on a motor learning task, a Rorschach-like perception problem, and a word association test.

Siipola (1950) and Siipola, Kuns, and Taylor (1950) experimented with the effect of color on Rorschach reactions. It was found that forms which are colored compatibly with the most frequently seen percept elicit rapid reaction times compared to incompatibly colored forms. Where color and form are incompatible, *S* is faced with competing response tendencies leading to longer reaction times. The Rorschach situation is thus conceived by Siipola as containing a "difficulty" continuum. A more difficult Rorschach problem is one which sets competing and incompatible response tendencies into operation. It is predicted that HA *S*s in comparison to LA *S*s respond to Rorschach stimuli with a greater difference in reaction time between compatibly (less difficult) and incompatibly (more difficult) colored blots. This effect will be augmented by stress instructions.

---

[1] This article is based upon a dissertation submitted to Yale University in partial fulfillment of the requirement for the degree of doctor of philosophy. The author would like to acknowledge S. B. Sarason's contribution and encouragement.

Similarly, data presented by Rapaport (1946, pp. 44, 55) indicate that stimulus words associated with longer response times tend to elicit comparatively more response words than stimulus words associated with relatively quick reaction times. These data suggest that the word association test contains a difficulty continuum. More difficult words are those which elicit relatively many different responses and are associated with longer reaction times. It is predicted that HA *S*s compared to LA *S*s will have a greater difference in reaction time between less difficult and more difficult words on the word association test. This effect will be increased by stress instructions.

Method

Four groups of 10 *S*s each were used. Two of these groups were extreme HA *S*s, and two were extreme LA *S*s as determined by the Sarason test anxiety scale. One group of HA *S*s and one group of LA *S*s were given task instructions designed to induce stress. The remaining two groups were given neutral instructions. All *S*s were presented with the following tasks in the order listed: (*a*) a motor learning task, (*b*) a free association word list, and (*c*) a Rorschach-like perceptual task.

Tasks

Motor learning experiment. A "bell maze" was constructed of eight wooden blocks mounted on a baseboard. Each of these wooden blocks had a different number of ordinary doorbell buzzers mounted on it. The number of buzzers on each block varied from two to nine. Each *S* was required to proceed from block to block and to learn the correct buzzer on each block. The correct buzzer was the one which rang when pressed. Thus, *S* was required to learn the correct path through a series of eight blocks. The problem contained three different correct paths, only one of which was activated at a given trial. After three practice trials, *S* was given 18 trials (six trials with each path) in which to learn the task. *S* was told which path he was to attempt before each trial. The number of errors made on each block on every trial was recorded.

Word association experiment. Twenty relatively "nonemotional" words were culled from Kent and Rosanoff's (1927) and Rapaport's word association lists. Ten of these words were extremely highly associated with a single response word (E words). Ten were extremely low with respect to this variable (D words). These two groups of words were matched for frequency in written language according to the Thorndike word count (1921). *S*s were required to state the first association which came to mind. Reaction time and the associations elicited were recorded. Data concerning the word list appear in Wiener (1954, Appendix B).

Perception experiment. Thirteen ink blots were used. Seven blots were colored compatibly with the most frequently elicited response (E blots). Six blots were colored incompatibly with the most frequently elicited response (D blots). Eleven of the 13 blots used were details taken from the standard Rorschach blots. Of these, nine were used by Siipola (1950). The blots, their source, and data concerning associated responses appear in Wiener (1954, Appendix B).

Instructions

*S*s given stress instructions were led to believe that their intelligence, abilities, and personalities were being intensively studied and that data obtained from this experiment would be compared with their school records. In the initially presented motor learning task, *S*s were told that impossible-to-attain scores were the norm for students who have an average adjustment. These false norms were reinstated during the course of the experiment (at the sixth and twelfth learning trials) in an attempt to increase stress as the experiment progressed. None of the stress *S*s attained the suggested norms.

Nonstress *S*s were told that the task materials were being pretested and that the results in no way bore any relation to their abilities. Before the motor learning task was presented, nonstress *S*s were correctly told that the problem was impossible to master and that they were not expected to do so, but that they should do their best.

Subjects

All 40 *S*s were male students in an introductory undergraduate course in psychology. They were given the Sarason test anxiety schedule about three months previously. None of the *S*s verbalized any suspected connection between the anxiety schedule and this experiment. HA Ss were in the top 10% of the distribution. LA *S*s were within the lower 10%. A total of 432 *S*s comprised the population of the test anxiety scores. *S*s were supplied to the experimenter so that he did not know whether a given *S* belonged to the HA or LA category.

Results

Motor learning experiment. It was predicted that anxiety level, stress instructions, and difficulty interact: HA *S*s compared to LA *S*s make more errors on the more difficult part of the problem than on the less difficult part. This effect should be greater under stress instructions than under nonstress instructions.

Table 1 indicates that there is no evidence that HA nonstress *S*s, compared to LA nonstress *S*s, make significantly more errors on D compared to E aspects of the task ($t < 1.$).[2] HA stress *S*s compared to LA stress *S*s are more impaired on the more difficult aspects of the test ($t = 2.08$, 18 df, $p = .03$). The interaction among anxiety level, stress instructions, and difficulty level is significant only at the .12 level ($t = 1.11$, 36 df).

Table 1 also indicates that LA stress *S*s compared to LA nonstress *S*s make fewer errors on the more difficult part of the task ($t = 3.69$, 18 df, $p < .01$). This finding considered alone would suggest that stressful situations are associated with heightened achievement motivation in LA *S*s. However, data from Table 1 show also that HA stress *S*s compared to HA nonstress *S*s make fewer errors on the more difficult part of the problem ($t = 3.23$, 18 df, $p < .01$, two-tailed test of significance). This unexpected finding may indicate that when all 18 learning trials are considered, the enhancing effect of increased motivation to do well obscures any impairing effect of the interfering responses.

In summary, HA stress *S*s compared to LA stress *S*s are impaired more as difficulty increases. This suggests that interfering anxiety-reducing responses hinder performances as a function of difficulty. LA stress *S*s compared to LA nonstress *S*s made fewer errors as difficulty increased. This implies that the effect of increased motivation, without corresponding interfering responses, is to perform better as difficulty increases. However, HA stress *S*s compared to HA nonstress *S*s were more efficient on difficult parts of the task. It is possible that HA stress *S*s experienced both impairing anxiety as well as heightened motivation to do well.

Perception experiment. It was predicted that instructions, *S*s' anxiety level, and difficulty level interact so that (*a*) HA stress *S*s compared to LA stress *S*s take longer in reacting to D than to E blots, and (*b*) this effect is greater under stress than under nonstress instructions. Examination of the D-E difference in Table 2 shows that this inter-

Table 1

Mean Errors on Simple (E) and Difficult (D) Aspects of the Motor Learning Problem

	Stress		Nonstress	
	E	D	E	D
HA	63.9	161.8	59.2	180.5
LA	49.5	122.2	52.8	156.1

---

[2] The statistical procedures used in testing hypothesis involving the motor learning task are outlined and discussed by Wiener (1954, Appendix D). Analysis of variance or mean difference statistics were not used because of scaling problems associated with different levels of difficulty.

Table 2

Mean Reaction Times in Seconds to E and D Ink Blots

	Stress		Nonstress	
	HA	LA	HA	LA
E blots	4.40	6.74	5.73	3.88
D blots	8.20	7.25	6.57	5.35
D−E Difference	3.80	.51	.84	1.37

action obtains in fact ($F = 5.03$, 1 and 36 *df*, $p = .02$).[3]

HA stress *S*s compared to LA stress *S*s have longer reaction times to D than to E blots ($t = 2.20$, 18 *df*, $p = .02$). Also, HA stress *S*s compared to HA nonstress *S*s have longer reaction times to D than E blots ($t = 2.48$, 18 *df*, $p = .01$).

Word association experiment. Differential reaction times to D and E words did not prove to be a useful response measure. None of the comparisons among groups approached significance, although all *S*s responded with greater latency to D than E words ($t = 3.70$, 39 *df*, $p < .01$). *S*'s anxiety level and stress instructions did not appear to affect differential reaction times. In view of this, the degree of the disturbance of the associated word was utilized to reflect impaired functioning. Using Rapaport's criteria and Janis's (1947) restatement of them, each *S* was scored for the percentage of disturbance on E and D words. All types of disturbances were grouped together. The data are summarized in Table 3.

HA *S*s compared to LA *S*s respond with more disturbed reactions to D compared with E words ($F = 4.96$, 1 and 36 *df*, $p = .02$). HA stress *S*s compared to LA stress *S*s have more disturbed associations to D words ($t = 2.89$, 18 *df*, $p = .01$). This effect does not obtain in nonstress situations ($t < 1$). The interaction between anxiety, instructions, and the E-D variable is not significant ($F < 1$). LA stress *S*s compared to LA nonstress *S*s tend nonsignificantly to be impaired less proceeding from E to D words in terms of where they respond with disturbed associations ($t = 1.40$, 18 *df*, $p = .09$).

Table 3

Mean Difference (D−E) Scores of Percentage of Word Association Disturbance

	Stress		Nonstress	
	HA	LA	HA	LA
D−E Disturbance Difference	22	5	22	15

When the content of the response word is considered, HA stress *S*s respond with more disturbed associations on D words, whereas LA stress *S*s tend not to respond differentially to E and D words.

Discussion

Responses to diagnostic projective tests may be in part a function of how individuals with different personality characteristics respond to variations of conflicting cues or difficulty levels within problem-solving situations. Test anxious *S*s experience anxiety specifically in a test situation—a situation in which achievement motives are elicited. This anxiety probably arises in part as a response to a previously learned conflict which is induced by the elicited achievement motivation. Therefore, it is assumed that in the test situation, stress instructions given to HA *S*s evoke an increment in the degree of anxiety these *S*s experience. HA *S*s, then, react to achievement-related conflicts with heightened anxiety, which leads to impairing anxiety-reducing responses.

It is postulated that HA *S*s are individuals who react to all test-related conflicts with impairing anxiety, whether the conflict is internal or external to the HA individual. Since HA stress *S*s react differently to E and D items, it is suggested that there is an interaction between *S*s' internal conflicts and the external conflict situation: HA stress *S*s approach a conflict situation with anxiety and therefore cannot effectively resolve the problems presented by a conflict situation. Thus, conflict situations engender anxiety in HA *S*s so as to impair performances, and performances in conflicting situations are also impaired as a function of *S*s' anxiety level. Child's interpretation seems very much similar to this: ". . . in complex situations, where the subject is already in conflict between various response tendencies relevant to the task, the presence of irrelevant responses made to anxiety heightens the conflict and interferes with performance . . ." (1954, p. 154).

---

[3] Wiener (1954, Appendix D) provides a rationale for utilizing analysis of variance statistics for projective tasks.

It is assumed that LA stress *S*s compared to LA nonstress *S*s have more achievement motivation and that both groups are relatively lacking in anxiety-produced interfering responses to test situations. Increased motivation, then, appears to be associated with less impairment as a function of response competition. Possibly, low motivation is associated with a task attitude of not attending to the more difficult aspects of the problem, whereas heightened motivation causes *S*s to attend to E and D items equally well.

The results suggest that HA stress *S*s and LA nonstress *S*s tend to manifest similar overt behavior. The tendency for these two groups to behave similarly is associated with different dynamic antecedents. HA stress *S*s are impeded by their anxiety from dealing effectively with the external problem situation. It appears that LA nonstress *S*s, because of their relative lack of involvement, are not motivated to deal effectively with an external problem situation.

It is possible that HA stress *S*s may have experienced increased motivation to do well along with impairing anxiety-reducing responses. If this were so, the facilitating effect of increased motivation might partially obscure the impairing effect produced by interfering responses.

Summary

This study was designed to test the effects of anxiety, stress instructions, and difficulty level on a motor learning task and to generalize the effects of these variables to a Rorschach-like perceptual task and a word association task.

Four groups of 10 *S*s each were used: a high anxious (HA) and a low test anxious (LA) group under stress instructions; HA and LA groups under nonstressful instructions. Each of these four groups was given a motor learning, word association, and perceptual task in the order listed. All experimental materials contained simple (E) and difficult (D) aspects. The D aspects contained a relatively greater amount of competing cues.

HA stress *S*s compared to LA stress *S*s were more impaired going from E and D items within the tasks. This finding was obtained for the motor learning, perception, and word association tasks. It is assumed that HA individuals under stress instructions experience interfering anxiety-reducing responses and these interfering responses interact with conflicting environmental cues to cause impaired performance.

Stress instructions appear to affect LA *S*s so that LA stress *S*s compared to LA nonstress *S*s tend not to be impaired going from E to D items. LA stress *S*s compared to LA nonstress *S*s make fewer errors on the D aspect of the motor learning task and tend not to perform differently to E and D words of the word association task in responding with disturbed associations.

This experiment's results suggest that the response competition definition of difficulty can be applied to projective materials. Projective tasks, in part, may be viewed as problem-solving situations.

Received June 11, 1958.

REFERENCES

CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–170.

FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.

JANIS, I. L. Psychological effects of electric convulsive treatments. Unpublished doctoral dissertation, Columbia Univer., 1947.

LAZARUS, R. S., DEESE, J., & OSLER, SONIA F. The effects of psychological stress on performance. *Psychol. Bull.*, 1952, **49**, 293–317.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

RAPAPORT, D. *Diagnostic psychological testing.* Vol. II. Chicago: Year Book Publishers, 1946.

ROSANOFF, A. J. *Manual of psychiatry.* (6th ed.) New York: Wiley, 1927. Ch. VII, pp. 546–620.

SARASON, S. B., MANDLER, G., & CRAIGHILL, P. G. The effect of differential instructions on anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 561–565.

SIIPOLA, ELSA M. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358–382.

SIIPOLA, ELSA M., KUNS, F., & TAYLOR, VIVIAN. Measurement of the individual's reaction to color in ink blots. *J. Pers.*, 1950, **18**, 153–171.

TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

THORNDIKE, E. L. *The teacher's word book.* New York: Teachers Coll., Columbia Univer., 1921.

WIENER, G. The interaction among anxiety level, stress instructions, and level of difficulty. Unpublished doctoral dissertation, Yale Univer., 1954.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

RELATIONSHIPS AMONG DIRECT AND INDIRECT MEASURES OF THE ACHIEVEMENT MOTIVE AND OVERT BEHAVIOR

DAVID MARLOWE
Ohio State University

Investigators planning studies which include n Achievement as one of the relevant variables can choose among several tests which purport to measure n Achievement. One of the most commonly used measures of n Achievement is McClelland, Atkinson, and Lowell's TAT-fantasy method (1953). The Personal Preference Schedule (PPS) recently published by Edwards (1954) also yields a score for n Achievement, a score which is more easily obtained in comparison to the fantasy method. Despite the great dissimilarity of these two methods, both have been used for the same purpose (predicting some behavior supposedly related to the strength of n Achievement). It seems necessary that some comparisons be made to determine whether these tests do in fact measure the same thing in the same way. McClelland (1958) has cited an unpublished study by Birney which reported a lack of any relationship ($r = -.002$, $N = 300$) between the TAT-fantasy and PPS measures of n Achievement. Additional data as to their utility for predicting overt achievement-oriented behavior would also be of some importance. In the absence of such information it seems inadvisable to make generalizations concerning achievement behavior when the data being compared were obtained by the different methods.

The purpose of the present study was to determine what relationships exist among three measures of achievement behavior. The three tests used were the TAT-fantasy method, the n Achievement scale of the PPS, and a sociometric measure. Previous studies have typically investigated the relationship of n Achievement scores to specific task performances in experimental settings. The sociometric measure which served as the criterion variable in the present study would seem to have the advantage of greater generality in terms of real-life behavior.

Rotter's Social Learning Theory (SLT) (1954, 1955) states that to maximize prediction from test responses it is necessary to differentiate among, and then evaluate systematically, the internal characteristics of the individual (his goal preferences and expectancies). The psychological situation must be treated as a third determinant of goal-directed behavior. Several studies dealing with achievement motivation offer empirical support for this conceptual scheme. French (1955), and in particular Atkinson (1953) and others (Atkinson & Raphelson, 1956; Atkinson & Reitman, 1956) have emphasized the need to take account of the achievement expectancies aroused by situational cues in order to maximize the prediction of overt achievement-oriented behavior from a measure of the strength of internal motivation.

Neither the fantasy measure nor the PPS makes any systematic use of the concepts of expectancy, goal preference, or the psychological situation. It seemed likely, therefore, that the fantasy and PPS measures of n Achievement would yield low correlations with a criterion measure of overt achievement behavior. In addition, SLT would lead one to expect that the fantasy and PPS measures would have a low correlation with each other. This latter assumption is based on the belief that the cues provided by a relatively unstructured fantasy situation (indirect measure) will lead the subjects (*S*s) to develop expectancies for satisfactions that differ in kind from those developed in a highly structured "self-report" situation (direct measure). Should an individual's expectancies in regard to two psychological tests differ greatly, then the scores obtained on the tests are not likely to be highly congruent.

Method

Subjects

Forty-four male undergraduate students comprising the entire membership of a social fraternity at Ohio State University participated in the study. Each *S* was a member of the fraternity for at least six months and a student at the University for at least one year. The cooperation of the *S*s was obtained by paying the fraternity for the approximately 55 minutes necessary for testing.

Procedure

Each *S* wrote six four-minute imaginative (TAT-type) stories under neutral group conditions according to the standard procedure of McClelland et al. (1953). Cards 2, 8, 1, 7, 28, and 4 of the standard series were administered in that order. The fantasy n Achievement scores were obtained by scoring the stories for Achievement Imagery and Achievement Thema according to the definitions of McClelland et al. (1953). The possible range of scores was 0 to 12. A mean of 5.5 and a standard deviation of 2.51 were obtained for the fantasy test. A measure of the interrater reliability for the fantasy test was obtained by correlating (Pearson r) the scores obtained by the writer with those independently obtained by another scorer. The resulting correlation, based on the stories of 20 *S*s, was .96.

Following the completion of the stories, the *S*s were given 55 pairs of items from the PPS, 28 of which comprised the n Achievement scale of the PPS. The remaining 27 items were used as buffers. The mean score for the n Achievement scale of the PPS was 12.4, with a standard deviation of 3.41. The present mean of 12.4 is significantly lower ($p <$ .001) than the mean reported by Edwards (15.6) for the males in his normative group of college students. This difference can be attributed to the highly selected nature of the present sample, which consisted entirely of fraternity members who were all enrolled in the College of Agriculture. The standard deviation of 3.41 does not differ significantly from the standard deviation of 4.13 reported by Edwards.

The *S*s were then administered a sociometric measure of achievement behavior which was adapted from Fitzgerald's (1958) measure of overt dependency. This instrument, which is based on a nominating technique, served as the criterion measure and yielded a quantified index of overt achievement behavior. Two descriptions were devised for the present study to describe the extremes of achievement behavior. The statements describing behaviors considered most representative of overt achievement-striving were:

Description A

Studies long and hard before exams
Regardless of the activity strives to be best
Likes to demonstrate his abilities and make a good impression
Sets very high standards for himself which he strives to attain

The statements describing behaviors considered least representative of overt achievement-striving were:

Description B

Finds reasons not to study before exams
Doesn't seem concerned whether he comes out first or last in competitive activities
Seldom gets involved in tasks requiring great skill or effort
Seldom strives to improve his performance

Each *S* was provided a copy of the descriptions which were labeled Description A and Description B, along with a list of the names of all the fraternity members. *S*s were then asked to select from the list of members "the name of the man who, in general, best fits Description A." The *S*s then selected the name of the man who "best fits Description B." Following this selection, the *S*s chose the man who "next best fits Description A," and then the man who "next best fits Description B." The nominations were thus made alternately, and in this manner each *S* made three nominations for each description. The nominations each *S* received on Description A were scored +3 for "best fits," +2 for "next best fits," and +1 for "third best fits." Similarly, the nominations for Description B were scored −3, −2, and −1. An *S*'s total score was the algebraic sum of all the nominations he received. A constant was added to each *S*'s total score to convert all the scores to positive

Table 1
Intercorrelations Among the Measures of Need Achievement and Overt Achievement Behavior
($N = 44$)

	PPS	Sociometric
TAT	−.05	.33*
PPS		.09

* $p < .05$ level.

values. The sociometric ratings yielded a corrected split-half reliability of .94. As in the Fitzgerald study, the method of nominating only the extremes of the group provided a normal distribution of scores.

Results

An analysis of the relationship between the fantasy and PPS measures of n Achievement and overt achievement behavior was made by correlating (Pearson r) the sociometric ratings with the scores on the fantasy and PPS tests. It is evident from Table 1 that there is no relationship between the PPS measure of the strength of n Achievement and the sociometric index of overt achievement behavior. There is a significant relationship ($r = .33$, $p < .05$) between the fantasy measure and the sociometric ratings.

The amount of relationship between the fantasy and PPS measures of n Achievement was evaluated by correlating (Pearson r) the scores on the fantasy and PPS tests. Table 1 indicates that the data obtained from these instruments are dissimilar and that the fantasy and PPS tests are not equivalent measures of n Achievement.

Discussion

According to SLT, the degree of relationship between a measure of need strength and overt behavior is typically of an unimpressive order because of the failure to take into consideration the view that any given situation has within it possibilities for the satisfaction of many different needs. Which particular need an individual will seek to satisfy will depend on the value of that satisfaction for him, and his expectancy as to whether his behavior will lead to that satisfaction in a particular situation. The absence of any relationship between the PPS measure of n Achievement and the sociometric ratings is therefore not unexpected in view of Edwards' failure to systematically distinguish between what an individual says he would like to do and what he says he has done in the past. It appears that the PPS is a confounded measure of n Achievement because of the failure to consistently word items in the same manner and differentiate between statements of expectancy and statements of goal preference. It can be concluded that the PPS is not a valid predictor of overt achievement behavior in a college population when peer ratings are used as the criterion measure.

The TAT-fantasy measure is significantly related to the sociometric ratings and appears to be a more valid predictor of overt achievement behavior than the PPS. It should be noted that reliabilities between .54 and .78 have been reported for a six-picture TAT measure of n Achievement (Haber & Alpert, 1958). This suggests that a more substantial, but still only moderate, relationship between the fantasy measure and overt behavior might have been obtained with a more reliable form of the fantasy test. For purposes of prediction, however, the obtained fallible scores yield a correlation of .33, which would result in only a 5.6% reduction in errors of prediction. The results of this and earlier studies (Atkinson, 1953; Atkinson & Raphelson, 1956; Atkinson & Reitman, 1956; French, 1955) clearly indicate the necessity of considering overt achievement-striving a complex, multiply determined behavior and suggest, further, that the failure to do so is likely to result, at best, in only low to moderate correlations between a measure of motive strength and some behavioral criterion. It is important to emphasize that the fantasy scores of n Achievement in the present study were based on a theme count. When "clinically" interpreted, the fantasy method might be a more valid predictor of achievement or other kinds of behavior.

The absence of any relationship between the fantasy and PPS measures indicates that these tests are not equivalent measures of n Achievement and that inferences from one test cannot be generalized to the other when n Achievement is the relevant variable. Thus, the present data support the earlier finding of Birney, and that of de Charms et al. (1955), who also found that *S*s perform in dissimilar manners on two tests which supposedly measure the same thing (n Achievement) when one test is an indirect measure of the motive and the other a direct measure. A possible explanation for these findings, on the basis of SLT, is that the difference in test performance may be mediated via the different satisfactions that the *S*s expect to obtain in the two situations. The fantasy situation may provide cues that lead to expectancies for satisfactions related to being creative and imaginative. In contrast, the self-report situation may lead to the development of expectancies for satisfactions related to being self-revealing or introspective.

Summary

The purpose of the present study was to determine the degree of relationship between two measures of n Achievement (the TAT-fantasy method and the n Achievement scale of the PPS) and a criterion measure of overt achievement behavior; and to determine the degree to which the two n Achievement measures yield equivalent scores. On the basis of Rotter's Social Learning Theory, it seemed likely that the two measures of n Achievement would yield low correlations with a sociometric index of overt achievement behavior. It seemed, further, that the two measures of n Achievement would have a low correlation with each other.

Forty-four male undergraduate subjects who were members of the same social fraternity were administered the TAT-fantasy and PPS tests of n Achievement, followed by the sociometric measure which was based on a nominating technique. The relationships among these instruments were analyzed by computing the product-moment correlations between the three measures.

The obtained correlations indicate that there is no relationship between the PPS and the sociometric measure, and no relationship between the PPS and the TAT-fantasy test. A significant correlation ($r = .33$) was obtained between the TAT-fantasy test of n Achievement and the sociometric measure. It was concluded that for a college population the fantasy test appears to be more suitable than the PPS for the prediction of overt achievement behavior.

The results were interpreted on the basis of Rotter's contention that additional variables are necessary if greater predictive utility is to be obtained with tests that measure the strength of internal motivation. The lack of relationship between the TAT-fantasy test and the PPS was interpreted as being due to the different kinds of expectancies which the *S*s develop in the fantasy and PPS situations. The correspondence between the theoretical interpretation of the present data and previous research findings was noted.

Received June 11, 1958.

REFERENCES

ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.

ATKINSON, J. W., & RAPHELSON, A. C. Individual differences in motivation and behavior in particular situations. *J. Pers.*, 1956, **24**, 349–363.

ATKINSON, J. W., & REITMAN, W. R. Performance as a function of motive strength and expectancy of goal-attainment. *J. abnorm. soc. Psychol.*, 1956, **53**, 361–366.

DE CHARMS, R. C., MORRISON, H. W., REITMAN, W. R., & MCCLELLAND, D. C. Behavioral correlates of directly and indirectly measured achievement motivation. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955. Pp. 414–423.

EDWARDS, A. L. *Manual. Personal Preference Schedule*. New York: Psychological Corp., 1954.

FITZGERALD, B. J. Some relationships among projective test, interview, and sociometric measures of dependent behavior. *J. abnorm. soc. Psychol.*, 1958, **56**, 199–203.

FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.

HABER, R. N., & ALPERT, R. The role of situation and picture cues in projective measurement of the achievement motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society: A method of assessment and study*. Princeton, N. J.: Van Nostrand, 1958. Pp. 644–663.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MCCLELLAND, D. C. Methods of measuring human motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society: A method of assessment and study*. Princeton, N. J.: Van Nostrand, 1958. Pp. 7–42.

ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

ROTTER, J. B. The role of the psychological situation in determining the direction of human behavior. In M. R. Jones (Ed.), *Nebraska symposium on motivation, 1955*. Lincoln: Univer. Nebraska Press, 1955.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

AN APPRAISAL OF TAULBEE AND SISSON'S "CONFIGURATIONAL ANALYSIS OF MMPI PROFILES OF PSYCHIATRIC GROUPS"

SOL L. GARFIELD AND JON SINEPS
University of Nebraska College of Medicine[1]

In a recent article in this journal, Taulbee and Sisson (1957) presented a technique of objective configurational analysis applied to the MMPI profiles of schizophrenic and neurotic patients. Sixteen scale pairs were utilized, which differentiated significantly two criteria and three validation groups of psychiatric patients. According to the authors, "The findings of this study indicate that the application of a configurational analysis technique for analyzing MMPI data yields an effective means for differentiating groups of neurotic and schizophrenic patients" (Taulbee & Sisson, 1957, p. 416).

In their study, Taulbee and Sisson presented specific cutoff scores for differentiating their schizophrenic and neurotic patients. According to their scheme, a scale pair score of 13 or more is a likely indicator of neurosis, while a score of 6 or less is considered indicative of schizophrenia. Any score between these two values is considered as "indeterminate," or nondiagnostic for these specific categories of patients. They also mentioned that "the lack of a normal sample and other psychiatric groups limits the application of these findings" (1957, p. 417). However, they felt that where there is a problem of differentiating between neurotic and schizophrenic conditions, "the cautious application of these scale pairs may be quite useful" (1957, p. 417). Some mention was made also that subsequent research indicated that latent schizophrenics, neurological cases, and normals frequently score in the "indeterminate" range between 7 and 12.

The Present Study

Since, in the clinical situation, most referrals are for a general personality and diagnostic appraisal rather than to differentiate between two specific categories such as neurosis and schizophrenia, the pragmatic test of the diagnostic value of any technique is to see how successive referrals are classified or diagnosed. In most instances the diagnostic possibilities for a given patient include other labels than just schizophrenia or neurosis. An effective technique for general clinical practice must differentiate not only between two selected groups of patients, but should be able to differentiate these groups from other diagnostic groups. Frequently a diagnostic technique which differentiates one pathological group from a control group fails when it is applied to the range of psychopathology encountered in the clinic or hospital. Consequently, we were interested in seeing how this particular approach works with a regular run of diagnostic referrals.

From our files, 129 case records were selected. These cases represented all cases of a nonorganic nature given the MMPI as part of their diagnostic test battery over the past two years and for whom there was a specific final diagnosis from the psychiatric staff. This group of patients was a heterogeneous group with diverse diagnoses. Sixteen were diagnosed as schizophrenic, 35 as neurotic, and the others were distributed among personality disorders, antisocial reaction, other psychotic conditions, etc. The mean age for the total group was 32.7 years (*SD* 13.4), with the means for the various subgroups being close to this figure. For example, the mean age for

---

[1] Nebraska Psychiatric Institute.

Table 1

MMPI Classification of Cases Diagnosed as Schizophrenia and Psychoneurosis

Patient Diagnoses	Sex	*N*	MMPI Classification		
			Correct	Incorrect	Indeterminate
Schizophrenia	M	9	6	1	2
	F	7	4	0	3
Psychoneurosis	M	12	1	5	6
	F	23	6	4	13
Total		51	17	10	24

the schizophrenic group was 31.1 (*SD* 12.8), while the mean for the neurotic group was 32.1 years (*SD* 9.9). The mean educational level of the total group was 11.1 grades in school (*SD* 2.3), and it was practically the same for all subgroups. The group consisted of 74 males and 55 females.[2] Since Taulbee and Sisson utilized only male subjects, separate analyses were made for each sex.

A configurational analysis for each case was made, following the stated criteria of Taulbee and Sisson. The cases were separated into two groups. All those patients for whom there was a final diagnosis of either psychoneurosis or schizophrenia were analyzed separately. These findings are presented in Table 1. The data on the other cases were also analyzed, and the results are given in Table 2.

As can be seen from Table 1, a majority of the patients diagnosed as schizophrenia are correctly classified by the MMPI scales. Only one case out of the 16 is misclassified, while five are categorized as "indeterminate." However, a different pattern of results was obtained for those patients diagnosed as psychoneurotic. In fact, in the latter group only seven patients out of 35 were correctly classified, the remainder being misclassified or in the indeterminate category. In terms of the data presented in Table 1, 33% of these cases

Table 2

MMPI Classification of Cases with Diagnoses Other Than Psychoneurosis and Schizophrenia

Patient Diagnoses	Sex	*N*	MMPI Classification		
			Schiz.	PN	Indeterminate
Antisocial reaction Sociopath	M	19	14	1	4
	F	2	1	0	1
Personality disorder Passive-Aggressive; Emot. unstable	M	17	10	2	5
	F	13	4	2	7
Other psychotic conditions	M	5	0	2	3
	F	5	3	1	1
Psychophysiological reactions	M	2	1	1	0
	F	3	1	1	1
Other Adjust. Diffic.; Alcoholism	M	10	5	0	5
	F	2	1	0	1
Total		78	40	10	28

---

[2] Comparisons were also made of the mean age and education for each sex in the various groups and did not reveal any unusual differences.

would be correctly diagnosed, almost 20% would be incorrectly classified, and the remainder would fall in the indeterminate category. On the basis of these two diagnostic groups alone, the profile analysis used here would appear to have some limitations as an aid in psychiatric diagnosis. Certainly, the current results are much less favorable than those reported by Taulbee and Sisson.

As mentioned earlier, however, when cases are referred for psychological evaluation, the diagnostic question is not usually limited to schizophrenia or psychoneurosis. Frequently, many possibilities are entertained, and a more practical clinical test of the effectiveness of this configurational approach is to see how it works with other clinical groups.

The data secured with other types of patients for whom the diagnostic picture is not always clear are presented in Table 2. The most interesting finding is that slightly over one-half of these 78 patients would be classified as schizophrenic in terms of the Taulbee-Sisson configurational method. This is particularly true of patients diagnosed as antisocial or sociopathic. A somewhat similar finding is noted with those patients falling in the broad category of personality disorders made up predominantly of individuals with passive-aggressive patterns and emotional instability. In fact, an analysis of Table 2 indicates that a schizophrenic diagnosis is much more frequent from the patient groups represented here than is any other. Certainly the psychoneurotic pattern seems to be much more infrequent. In our total group of patients, only 18 received a clear-cut psychoneurotic score, whereas 59 were classified as schizophrenic. On the basis of these data, one might anticipate securing "schizophrenic" scores quite frequently with a broad variety of patient groups.

Discussion

On the basis of the data collected in this study, it appears that the profile method of Taulbee and Sisson may be less valuable in differentiating between schizophrenics and neurotics than previously reported. It also does not appear to offer much promise as a diagnostic measure when applied to the wide range of cases seen in clinical practice. With the exception of the small group of schizophrenics in this study, the psychoneurotic and other patient groups provide many false positives as well as a large number of indeterminate diagnoses. The false positives seem largely to occur with the scales designated as indicating a schizophrenic diagnosis. It would thus appear that many more incorrect and indeterminate diagnoses would be secured than would accurate ones.

Obviously, the same problems of comparability of sample populations and the reliability of psychiatric diagnosis apply to the current researches as to other studies which are based on groups differentiated in terms of such diagnostic criteria (Garfield, 1949). If such difficulties contribute to the problems of replicating the work of other investigators, as seems reasonable, it is apparent that the test patterns secured in one setting can only be applied with extreme caution to other settings.

Summary

The configurational analysis of MMPI profiles devised by Taulbee and Sisson was applied to a group of 129 patients in different diagnostic categories. In terms of the data obtained, it appears that this method offers little promise as a diagnostic measure when applied to unselected samples of patients in clinical practice. The perennial problems of the comparability of patient samples and the reliability of psychiatric diagnoses also are pertinent here.

Received June 16, 1958.

REFERENCES

Garfield, S. L. An evaluation of Wechsler-Bellevue patterns in schizophrenia. *J. consult. Psychol.*, 1949, **13**, 279–287.

Taulbee, E. S., & Sisson, B. C. Configurational analysis of MMPI profiles of psychiatric groups. *J. consult. Psychol.*, 1957, **21**, 413–417.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

THE EFFECTS OF WARM AND COLD INTERACTION ON THE ADMINISTRATION AND SCORING OF AN INTELLIGENCE TEST[1]

JOSEPH MASLING
Syracuse University

Many recent studies have examined the influence of the psychologist-subject relationship in projective testing. Despite the methodological errors (Hammond, 1954; Levy, 1956) which have characterized some of this research, there is evidence that subjects do not give the same responses to one examiner that they give to another, because of the instructions (Henry & Rotter, 1956), the reinforcement given the responses (Wickes, 1956), the situation (Klatskin, 1952), and the personality of the examiner (Lord, 1950; Sanders & Cleveland, 1953). Not only does the examiner influence the subject's test behavior, but it has been demonstrated (Masling, 1956) that a subject by acting "warm" or "cold" can influence an examiner's interpretation of responses to a projective test.

The ambiguity which faces both the examiner and the subject in a projective testing situation makes it probable that each will be influenced by the other in attempting to complete their respective tasks. In intelligence testing, however, the instructions are specific, the stimuli are clearly defined, and there are right and wrong answers. The examiner who administers an intelligence test is required to read the questions as stated in the text and to evaluate the answers with the aid of the scoring manual.

The examiner-subject relationship in intelligence testing has not received a great deal of attention. During the course of their training most examiners are exhorted to establish "rapport" and admonished to be "objective." The "objective" examiner is charged with the responsibility of deriving as valid an estimate of the intelligence of the subject as can be obtained, without regard for his personal attitudes about the subject. He is thus expected to be standardized and depersonalized. The purpose of the present study was to investigate the extent to which an examiner could divest himself of personal bias in administering and scoring an intelligence test—in this case, three verbal subtests of the Wechsler-Bellevue II (hereafter called the W-B II)—when the subject acted in either a highly approving, interested (warm) manner or in a persistently rejecting, disinterested (cold) manner. The specific hypotheses which were tested were as follows:

I. When an examiner tests two subjects, one of whom acts warm to him and the other cold, he will be more generous in scoring the responses of the warm subject.

II. During the course of administration of an intelligence test to two subjects, one of whom is warm and the other cold, an examiner will: (*a*) make more reinforcing statements to the warm subject than the cold; (*b*) ask more questions of the warm subject, giving him the opportunity of clarifying or reformulating an answer.

Method

The interaction. Manipulation of the interaction was effected through the use of attractive female accomplices, posing as test subjects (*Ss*), who acted either warm or cold to the examiner. In the warm condition, the accomplice acted interested in the examiner and in the test; she responded freely to his questions and tried to communicate respect and liking for him. In the cold condition, the accomplice acted disinterested and bored with the test and the examiner; her attitude was that of fulfilling an unpleasant class assignment which she wanted to complete as soon as possible. She tended to answer pretest interview questions in monosyllables and throughout avoided eye contact with the examiner. In the middle of each cold session the accomplice in a deliberate, calculated fashion put on sunglasses, thereby increasing the psychological distance between herself and the examiner.

---

[1] The author is greatly indebted to James Beaber, H. O. Beldin, Dan Briggs, Robert Gravely, John Henderson, Randall Martin, John McKinney, Louis Patulo, Eugene Quarrick, Donald Smith, and Gerald Studebaker who so willingly served as examiners; to Neila Dunay and Eileen Shapiro who performed competently as accomplices; and to Bernard Braen, Sanford Dean, Howard Friedman, Sidney Orgel, Bertram Rothschild, Jerome Schiller, Edward Siegel, and Herbert Tothill who served as judges in various parts of the study.

Examiners. The examiners were 11 graduate students at Syracuse University, four in the area of clinical psychology, three each in special education and remedial reading, and one in the area of developmental psychology. Each examiner had completed at least one course in the administration of individual tests of intelligence and, in addition, six were currently enrolled in or had completed at least one course in projective testing. The estimated number of W-Bs each had given previous to the study ranged from 18 to over 200, with a median of 21.

When each examiner was asked to participate in the study he was told that the author was interested in the comparability of various short forms of the W-B. He was informed that he would administer either two or three subtests and that the choice of particular subtests for each examiner would be made on a random basis. The *Ss* would be two undergraduates who were completing their assignment of participating in psychological experiments as part of the requirement of the introductory course in psychology. The examiners were asked not to communicate with each other regarding any phase of the study, and they were requested to urge the *Ss* not to discuss the study with their friends.

Testing session. Each examiner administered three subtests—Information, Comprehension, and Similarities—to both *Ss*, one of whom acted warm to him and the other cold. Each accomplice had five cold and five warm roles, with the order of presentation of conditions and accomplice counterbalanced.

Scoring. To insure uniformity of scoring procedures the examiners were requested to use as a guide to scoring only the instructions found in Wechsler's (1944) "Measurement of Adult Intelligence," third edition. They were asked not to use any supplementary manuals or monographs.

Test responses. The experimenter prepared a script for each accomplice to memorize and repeat to each examiner, regardless of the interaction condition she was to establish with him. For each accomplice, 14 of these responses were written to maximize difficulty in scoring. As examples of the answers that were used, Accomplice A's script called for her to make these responses:

Information subtest:

Item, population: "180 million"

Item, Faust: "It is an opera. I've listened to it, heard it. I'm not sure of the composer's name. It is French, or maybe German, something like G-o-n-o-t, Gonot. I really don't know." (She pronounced the name Gó-no.)

Item, Habeas Corpus: "Like a writ of Habeas Corpus? That gets you out of jail, when you're in, sometimes."

Similarities subtest:

Item, praise and punishment: "They are alike in that they are both ways of criticizing other people in order to improve their behavior."

Accomplice B's script contained these responses:

Information subtest:

Item, population: "120 million"

Item, Koran: "That is the bible of the Arab peoples."

Comprehension subtest:

Item, laws: "Laws establish a set of rules so that everyone knows what is expected of him."

Item, land: "Because it is more desirable."

Each accomplice was warned of the necessity of adhering rigorously to the script assigned her without making any substitutions or improvisations. If asked for additional information they were told to repeat the substance of the original answer, unless the script provided for a clarification of the response. An analysis of the typescripts showed that the accomplices did, in fact, repeat the assigned answers for each of the examiners, with but a few minor deviations from the script.

The testing room was equipped with a tape recorder and each examiner was asked to record both sessions. Examiner 9, however, had the input volume turned so low that it was impossible to obtain a typescript from either of his tapes, and Examiner 10 was unable to work the recorder for his first session. For this reason Hypotheses IIa and IIb were tested using the typescripts of the remaining eight examiners. Since all examiners turned in scored test forms, Hypothesis I was tested with an *N* of 10.

Results

There was no mistaking the impact of the warm and cold conditions on the examiners. All reported that one *S* seemed particularly disinterested in the test, with some emphasizing the notion that this represented "sick" behavior. One examiner correctly guessed that the *S*s were really accomplices, and he was therefore replaced and his data were not used. Krasner (1958) has reported that in studies attempting to establish conditioning of verbal behavior without awareness approximately 5% of the *S*s learn what the experimenter is trying to do. It is generally accepted practice not to use the data on these *S*s.

Hypothesis I was tested by comparing the way in which each examiner scored the responses given him under the two conditions. Since the experimenter had intentionally written responses that gave a higher "true" IQ for one of the accomplices than for the other, the bias of examiner scoring was determined from the mean of the 10 scores given each accomplice, rather than from the raw scores. Once a mean score for each accomplice had been obtained, the extent and direction of differences from each mean were derived for each examiner. To test for the probability of obtaining a distribution of differences such as this by chance, the Walsh test, as suggested by Siegel (1956), was used, yielding a one-tailed p value of .056.

Hypothesis I can also be evaluated by inspection of Table 1. Of the five examiners who tested Accomplice A under the warm condition, four gave her scores greater than the mean, while of the five examiners who tested her in the cold condition, four gave her scores smaller than the mean. The identical results were obtained for Accomplice B: four of the five examiners who interacted with her in a warm manner gave her scores greater than the mean, while four of the five examiners who interacted with her in the cold condition gave her scores smaller than the mean.

In order to test Hypotheses II, typescripts were made from the tapes of each interview. The eight pairs of tapes that were audible were transcribed and coded, with identification of the name of examiner, *S*, and type of interaction removed. Two judges were given typescripts that began with the first question on the Information subtest, the pre- and post-test interviews having been removed, and they

Table 1
Raw Scores (and Deviations from the Mean) Given Each Accomplice by Examiner and Interaction Condition

Examiner	W-B Previously Given	Accomplice A[a]		Accomplice B[b]	
		Warm	Cold	Warm	Cold
1	(200+)	47 (+2.7)			48 (−0.5)
2	(40)		43 (−1.3)	47 (−1.5)	
3	(50)	45 (+0.7)			45 (−3.5)
4	(20)		47 (+2.7)	49 (+0.5)	
5	(11)	41 (−3.3)			47 (−1.5)
6	(20)	46 (+1.7)			49 (+0.5)
7	(22)		43 (−1.3)	49 (+0.5)	
8	(20)		43 (−1.3)	52 (+3.5)	
9	(100+)	45 (+0.7)			48 (−0.5)
10	(18)		43 (−1.3)	51 (+2.5)	

[a] Mean score = 44.3.
[b] Mean score = 48.5.

Table 2

Examiner Remarks by Interaction Condition and Accomplice

Examiner	W-B Previously Given	Accomplice A				Accomplice B			
		Warm		Cold		Warm		Cold	
		R[a]	Q[b]	R[a]	Q[b]	R[a]	Q[b]	R[a]	Q[b]
1	(200+)	4	5					0	5
2	(40)			6	9	6	9		
3	(50)	9	10					8	8
4	(20)			24	10	23	8		
5	(11)	24	10					16	6
6	(20)	3	12					2	5
7	(22)			19	3	37	5		
8	(20)			12	2	17	5		

[a] Reinforcing remarks.
[b] Questioning remarks.

assigned every examiner remark to one of four categories: reinforcing statement, punishing statement, questioning statement, and miscellaneous. A reinforcing statement was defined as any remark designed to encourage, support, or reassure the *S* or the response he made ("OK," "swell," "I knew you could answer that," etc.). A punishing statement was defined as any remark designed to insult or disparage the *S* or the response he made ("I thought everyone knew that answer," etc.). A question was defined as any attempt made to have the *S* give additional information or to clarify an answer ("Tell me more about that," "Can you be more specific," etc.). The miscellaneous category was used for all other examiner remarks, including such phrases as "Uh-huh" and "Mm-hmm."

There were 285 examiner remarks in the testing sections of the 16 interviews. The judges independently agreed on the rating of 254 (89%) of them, the categorization of the 31 remaining remarks taking place by conference. The distribution of reinforcing and questioning remarks by examiner, accomplice, and condition is found in Table 2. Since the judges agreed that only two punishing remarks were made, no further work was done with this category nor was there any analysis of the wastebasket miscellaneous category.

Once the number of reinforcing and questioning statements had been obtained, a comparison was made of each examiner's verbal behavior during the warm interaction with his verbal behavior during the cold interaction by use of the Wilcoxon matched-pairs signed-ranks test (Siegel, 1956). Using one-tailed probability values for all tests of significance, it was found that examiners made more reinforcing statements to the warm *Ss* (p value between .01 and .025), and they asked more questions of the warm *Ss* ($p = .025$). The sum of reinforcing and questioning statements was also greater for the warm condition than the cold (p between .01 and .025).

While the experimental hypotheses were substantiated, the differences between conditions seemed much smaller than the differences among examiners. For example, Examiner 1 made a total of only 14 remarks to his *Ss*, while Examiner 7 made 64. Examiner 8's scoring favored the warm condition by 4.8 points, while Examiner 5's scoring was biased in favor of the cold condition by 1.8 points. Since it was possible that the more experienced examiners were least biased by the interaction, rank-order correlations were computed between the number of W-Bs previously given and the dependent variables. None of the rho's were significantly greater than zero.

Discussion

The results of this study indicate that the examiner-subject interaction influenced the psychologist's behavior in the administration and scoring of the three subtests of the W-B II. When the instructions to the cold accomplice are considered, i.e., to answer in monosyllables, to appear disinterested, and the typescripts of the sessions studied, it becomes clear that the examiners tried to make contact with the cold *S* and, in failing to do this, became silent. While the examiners were undoubtedly trained to encourage the *S*, this was difficult to do when friendly overtures elicited disinterest and rejection. The feelings which the interaction aroused in these examiners obviously influenced the manner in which they administered the tests. With warm, responsive *S*s they tended to encourage and question; with cold *S*s they tended to remain silent.

It is difficult to predict the extent to which this particular finding can be generalized to nonlaboratory situations. Probably few individuals taking intelligence tests act as hostile and nonparticipating as the cold accomplice. However, some Ss, notably children, may become threatened by the testing situation, responding with belligerence or silence or other variations of avoidance. Schafer (1954) has described the dynamics of adult *S*s who may not interact favorably with the examiner. Thus, the cold accomplice represents a caricature of certain kinds of *S* behavior, perhaps overdone, but posing problems which examiners face in milder forms with bona fide *S*s.

What effect does the style of administration have on the score earned by the *S*? Since the accomplices repeated the scripts prepared for them, no direct evidence is available from the present study. One experiment (Gordon & Durea, 1948) found, however, that a group given the Stanford-Binet under conditions of discouragement ("You're not doing well") earned a mean IQ score 6.35 points lower than a control group. Lantz (1945) demonstrated that the experience of success or failure may influence performance on intelligence test items. Hutt (1947) concluded that adaptive testing with the Stanford-Binet produced higher scores for poorly adjusted children than the traditional administration of the test. It is very likely, then, that *S*s who experience discouragement or frustration or who anticipate failure while taking an intelligence test may do less well than they might under other conditions of interaction. Since poorly adjusted, easily threatened *S*s may unwittingly discourage the examiner from relating to them favorably, their performance would suffer, not because of lack of potential but because of the interaction. This, along with their greater anxiety and distractibility, could account for the depressed IQ scores ordinarily found in poorly adjusted children.

The interaction also affected the examiners' "objective" judgment of the scoring of relatively "objective" material. Even though they had the Wechsler manual available, a response given in the warm condition tended to be given greater credit than the identical response given in the cold condition. This bias is even more striking when it is considered that the scoring occurred some time after the testing, allowing the examiners some perspective regarding the events of the session. Again, this study exaggerated the situation found in most clinic settings, since the examiners were given responses that were selected because they were difficult to evaluate. However, an examination of the scoring records indicated that there were systematic differences in scoring even for those responses which were cited as examples in the scoring manual.

The artificial nature of this study—the use of accomplices and relatively unsophisticated examiners, the exaggerated nature of the interaction, the use of ambiguous responses—together with the inadequate sampling of both the examiner and accomplice populations (Hammond, 1954) limits severely the generalization of these findings to nonlaboratory settings of psychologists and *S*s. What has been demonstrated is that in giving an intelligence test under these conditions an advanced graduate student examiner will respond to the way *S*s interact with him and will act out in administration and scoring his feelings about the interpersonal situation.

Summary

1. Eleven graduate students, each of whom had completed at least one course in the administration of individual intelligence tests, administered the Information, Comprehension, and Similarities subtests of the Wechsler-Bellevue II to two subjects. The test subjects were accomplices who acted in either a warm or cold role to the examiners, giving as their

responses memorized answers, 14 of which were specifically devised to be difficult to score. One examiner became aware of the purposes of the experiment and his data were not used. Each accomplice had five cold and five warm roles, and each examiner saw one subject who acted warm and one who acted cold.

2. Each testing session was taped, although only eight examiners were able to record audible tapes for both of their sessions. From the typescripts prepared from these 16 tapes, every examiner remark during the course of the testing part of the interview was assigned one of the following categories: reinforcing remark, punishing remark, questioning remark, miscellaneous. Of the 285 examiner statements, two judges independently agreed on the rating of 254 of them, for an agreement of 89%.

3. The results indicated that in scoring the responses, the examiners tended to be more lenient to the warm subject than the cold ($p = .056$). The examiners also tended to use more reinforcing comments ($p = < .025$ $> .01$), and to give more opportunity to clarify or correct responses ($p = .025$) to the warm subject. The magnitude of the differences in behavior to the two subjects was generally small, with individual differences more marked than differences due to the effect of the interaction.

Received June 27, 1958.

REFERENCES

Gordon, L. V., & Durea, M. A. The effect of discouragement on the revised Stanford-Binet scale. *J. genet. Psychol.*, 1948, **73**, 201–207.

Hammond, K. R. Representative vs. systematic design in clinical psychology. *Psychol. Bull.*, 1954, **51**, 150–159.

Henry, Edith, & Rotter, J. B. Situational influences on Rorschach responses. *J. consult. Psychol.*, 1956, **20**, 457–462.

Hutt, M. L. "Consecutive" and "adaptive" testing with the revised Stanford-Binet. *J. consult. Psychol.*, 1947, **11**, 93–104.

Klatskin, E. H. An analysis of the effect of the test situation upon the Rorschach record: Formal scoring characteristics. *J. proj. Tech.*, 1952, **16**, 193–199.

Krasner, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

Lantz, B. Some dynamic aspects of success and failure. *Psychol. Monogr.*, 1945, **59**, 6–21.

Levy, L. H. A note on research methodology used in testing for examiner influence in clinical test performance. *J. consult. Psychol.*, 1956, **20**, 286.

Lord, Edith. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, **64**, No. 10 (Whole No. 316).

Masling, J. The effects of warm and cold interaction on the interpretation of a projective protocol. *J. proj. Tech.*, 1957, **21**, 377–383.

Sanders, R., & Cleveland, S. E. The relationship between certain examiner personality variables and subjects' Rorschach scores. *J. proj. Tech.*, 1953, **17**, 34–50.

Schafer, R. *Psychoanalytic interpretations in Rorschach testing.* New York: Grune & Stratton, 1954.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Wechsler, D. *Measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.

Wickes, T. A. Examiner influence in a testing situation. *J. consult. Psychol.*, 1956, **20**, 23–26.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

MORE CONSTRUCT VALIDATION OF THE EGO–STRENGTH SCALE[1]

IRVING I. GOTTESMAN
University of Minnesota

The ego-strength construct has a high frequency in the conversations and reports of clinical psychologists. Its origins in psychoanalytic theory have not deterred the toughminded from its use; indeed, it was in the camp of empiricism that a scale purporting to measure it was derived. Surprisingly, Barron's (1953) new scale did not lead to an avalanche of original research either using it or refuting it. Barron (1956) has continued to work with the scale, Wirt (1955) has done a further validation of the prognostic utility of the Ego-Strength (*Es*) scale, and Tamkin (1957) and Tamkin and Klett (1957) have concerned themselves with its construct validity (Cronbach & Meehl, 1955). Within the framework of construct validation, tests of a number of hypotheses about ego strength are reported in this paper. *Es* scores on samples of adult and adolescent normal and psychiatric subjects were used. In conjunction with the data of other researchers, it is intended to form enough strands in the *Es* scale network to more nearly outline its meaning and utility.

Origins of the Ego Strength Construct

As early as 1894, Freud (1924) used the construct of *ego* but did not expound upon it rigorously until the publication of *Das Ich Und Das Es* in 1923 (Freud, 1927). His now classical topographical conception of mental life with its tripartite division was first made there. Inasmuch as the topic of consideration is ego strength and not ego, this detour must be defended on the grounds that an absolute concept of strength is without utility; strength of any kind is relative to the current and potential demands put upon it. A complete catalogue of ego functions would be extremely lengthy and partly idiosyncratic. Only some of the more important will be presented here. In Freud's definitive paper he tells us that "the ego has the task of bringing the influence of the external world to bear upon the id and its tendencies, and endeavors to substitute the reality principle for the pleasure principle which reigns supreme in the id . . . the testing of reality is rather one of the functions of the ego" (Freud, 1927, p. 30). Further teasing out of functions has been partly obviated by the legatees of the ego, the ego psychologists, including A. Freud (1946), Hartmann (1950), and Nunberg (1948). Hartmann captures the difficulty of the task by first defining ego negatively and then concluding that it is defined by its functions: "Ego is not synonomous with personality, or the individual; it does not coincide with the subject as opposed to the object of experience; and it is by no means only awareness of the feeling of one's own self" (p. 75). The ego organizes and controls motility and perception (of self as well as the world); it serves as a protective barrier against excessive stimuli; action and thinking are also ego functions; it anticipates, synthesizes, and inhibits. Any combination of these functions may occur simultaneously. The ego must defend itself against id impulses, the reproaches of a punishing superego, and the traumata of the real world.

---

[1] The author is indebted to the people who made the data for this study available and accessible. R. D. Wirt was not only the mentor for this paper but also cleared the path to data sources and furnished the MMPIs for the schizophrenic sample. P. F. Briggs obtained access to the Hathaway-Monachesi Study raw data, and M. Dunnette made a sample of salesmanagers available at the Minnesota Mining and Manufacturing Co. The A. H. Wilder Child Guidance Clinic, St. Paul, where the author is a fellow in psychology, was the source for the sample of emotionally disturbed adolescents.

Implicit in the above is a source and a quantity of energy to carry out the functions. In analytic contexts the energy is explicitly called libido; it in turn emanates from the id and is posited as instinctual energy in its original form. The available ego libido, in this theory, consists of the id libido that has been desexualized and deaggressivized, i.e., neutralized (Hartmann, 1950). It follows that any reference to ego strength must, in this context, pertain to the amount of available libido at any given time. Any instrument which attempts to measure this construct must be either a sample or a sign of this available energy. The enormity of this task is evident since the construct is in constant flux. The best that might be aimed for are situational confidence limits in its estimation. An ideal evaluation would require a balance sheet of psychic energy listing the intrapsychic and interpersonal assets, liabilities, and reserves. Living in a Korean foxhole would be an environmental liability, for example, and having superior intelligence would be an intrapsychic asset.

Barron's (1953) conceptualization of ego strength, derived from the *Es* scale item content and personality and intelligence test correlates, involves physiological stability and health, a strong sense of reality, feelings of personal adequacy and vitality, permissive morality, lack of ethnic prejudice, emotional outgoingness and spontaneity, and intelligence. In the light of the previous discussion his sampling of ego functioning may not be comprehensive enough for content validity.

Hypotheses

If the *Es* scale really measures ego strength, it should be able to discriminate between samples about whom relative ego-strength inferences can be made. Normal adults should have higher *Es* scores than normal adolescents. Normal adolescents should be higher than emotionally disturbed adolescents. Severe delinquent adolescents should be between the former and the latter. Normal adults should be higher than psychoneurotic and psychotic adults, and the second, higher than the third. Superior adults should have higher *Es* scores than any of the above groups. Ego strength has a positive linear relationship to personality integration so that *Es* scores should predict the same way. The item content of the scale suggested that a fake-good test-taking attitude could contribute to plus getting; therefore, it was hypothesized that *Es* was highly correlated with the *K* scale.

Procedure

Samples to whom the Minnesota Multiphasic Personality Inventory (MMPI) had been administered and about whom relative ego-strength inferences could be made were gathered. The tests were rescored for *Es* and *K*. All possible *t* tests were made in order to rank the samples. A rank order correlation was computed between the group means for *Es* and *K* and for one of the subsamples.

Samples

Normal adults. Hathaway and Briggs (1957) have provided normative data on some new MMPI scales for a sample almost identical with the original normative group. Their summary statistics for over 500 males and females on the two scales were used as reference.

Normal adolescents. A random sample of 31 boys was drawn from all the ninth-grade boys in the Minneapolis public schools who had no record of delinquent behavior. The raw data from the Hathaway-Monachesi study (1953) were the source of the sample.

Severe delinquent adolescents. A random sample of 31 boys was also selected from the above study, all of whom had been involved in serious delinquencies both before and after taking the MMPI.

Emotionally disturbed adolescents. An ongoing sample of 31 adolescents who were re-

Table 1

Es and *K* Scores for Study Samples

Group	*N*	*Es*	*SD*	*K*	*SD*
Superior adult	21	53.33	4.32	19.90	4.07
Severe delinquent	31	47.98	4.56	13.35	3.90
Normal adolescent	31	46.19	6.00	15.00	4.36
Hs-Hy profile	21	40.95	7.23	17.90	5.33
Schizophrenic	25	39.56	8.78	15.12	6.39
Disturbed adolescent	31	37.90	7.54	12.55	4.71
D-Pt profile	25	37.60	6.16	13.52	3.03

Table 2

Es Data from Other Studies

Group	*N*	*Es*	*SD*	Investigator
Improved VA hosp. patient	70	55.2	7.00	Wirt
Air Force officer	160	52.73	4.05	Barron
Graduate student	40	50.92	5.62	Barron
Greatly improved clinic patient	9	49.66	—	Barron
Greatly improved VA hosp. patient	19	48.6	7.03	Wirt
Normal male	226	44.33	6.21	Hathaway & Briggs
Improved clinic patient	27	43.07	—	Barron
VA mental hygiene clinic patient	52	41.79	7.38	Barron
Normal female	315	40.21	6.36	Hathaway & Briggs
Unimproved VA hosp. patient	114	33.3	6.96	Wirt
Unimproved clinic patient	16	32.75	—	Barron

ferred to the A. H. Wilder Child Guidance Clinic for diagnostic study was used. The sample was about equally divided between boys and girls with a median age of 14.5 years, range, 13 to 18 years. Although all were outpatients, some were severe neurotics or psychotic.

Superior adults. A sample of MMPIs of applicants for sales positions who were employed and subsequently were promoted to salesmanagers at the Minnesota Mining and Manufacturing Co. was selected. A search of the files for the tests of the managers who had been hired during the 10-year period in which the MMPI was used in selection revealed only 16 answer sheets. Five additional tests were selected randomly from among those of applicants who had successfully met the hiring requirements. The total of 21 men were superior to the norm group (Hathaway & Briggs, 1957), at least in education and occupational level. The representativeness of this small sample on *Es* was confirmed by the median *Es* score of a sample of 68 similarly employed men at another large Twin City corporation.[2]

Veterans Administration Hospital D *and* Pt *profiles*. The MMPIs of all patients admitted to the psychiatry service at the Minneapolis Veterans Administration Hospital in 1956 were examined. All tests that had scales *D* and *Pt* coded high (Welsh, 1948) in that order were selected for study. The profile and not the diagnosis was the criterion for selection. An *N* of 25 resulted.

Veterans Administration Hospital Hs *and/or* Hy *profiles*. The above admission MMPIs yielded 21 tests with scales *Hs* and/or *Hy* coded high. Again the profile was the criterion for selection.

Veterans Administration Hospital schizophrenics. The admission MMPIs on a group of Veterans Administration Hospital patients subsequently diagnosed as schizophrenic were selected for study. The sample consisted of 21 paranoid schizophrenics and 4 diagnosed schizophrenic reaction undifferentiated or unclassified. Analysis of the data showed no reason to handle them by subclass. The diagnosis of schizophrenia was the criterion for this sample of 25.

Results

Table 1 shows the results for *Es* and *K* for the seven samples. Table 2 shows some of the data collected on *Es* in other studies.

A series of two-tailed *t* tests for the *Es* scores gave results which may be displayed as follows. Groups which are connected by either an overscored or underscored line do not differ significantly from each other at the .05 level.

$$\begin{matrix}\text{Superior} \\ \text{Adult}\end{matrix}\quad \overline{\begin{matrix}\text{Severe} \\ \text{Delinquent}\end{matrix}\quad \underline{\begin{matrix}\text{Normal} \\ \text{Adolescent}\end{matrix}}}\underline{\quad \begin{matrix}\text{Normal} \\ \text{Male}\end{matrix}}\qquad \overline{\text{Hs-Hy}\quad \text{Sc}\quad \text{D-Pt}\quad \begin{matrix}\text{Disturbed} \\ \text{Adolescent}\end{matrix}}$$

---

[2] Rosen, Ephraim. Personal communication, May 8, 1958.

The normative group of males did not differ significantly from the normal adolescents. Severe delinquents were significantly higher than the normal males. All of the psychiatric groups were significantly lower than the nonpsychiatric groups. Some error has been left in the disturbed adolescent comparisons by not separating boys and girls; the normative data showed a raw score difference between sexes of four points in favor of males.

A series of *t* tests for the *K* scores of these groups gave these results:

Superior Adult	Hs-Hy	Sc	Normal Adolescent	D-Pt	Normal Male	Severe Delinquent	Disturbed Adolescent

A rank order correlation between *Es* and *K* for the seven study samples was computed. Rho equalled .46; with these few pairs one of .78 was needed for significance at the .05 level. Another rho was computed for the same variables for the subsample of 16 salesmanagers, and a correlation of .49 was obtained, which was significant beyond the .05 level. Barron (1953) obtained *r*'s of .31 for his samples of clinic patients and graduate students.

Discussion

A parsimonious interpretation of the results appears to be that the *Es* scale broadly discriminated between psychiatric and nonpsychiatric adults and adolescents. It did not discriminate different degrees of psychiatric incapacitation. It did not discriminate delinquent adolescents from nondelinquents, and predicted in the opposite direction for the former compared to normal adults. It showed no relation to age within the limits of this study. Its previously demonstrated utility as a predictor of favorable response to psychotherapy may be questioned, since severe delinquents, who are generally considered poor bets for psychotherapy, did not differ significantly from patients rated greatly improved.

The correlation between *Es* and *K* was higher than that originally reported and suggested some interpretations of what *Es* may be picking up. It may reflect a defensive test-taking attitude and the ability to recognize socially desirable descriptions of personality. Both of these behaviors show contact with and testing of reality. Inasmuch as ego defense and reality testing are both ego functions, the scale would measure these two aspects of ego strength. It would, however, only indicate strength if the defense were not carried to an extreme. Reacting to the threat of not being selected for a job and reacting to a threat from persecutory delusions would both contribute to plus getting on the *Es* scale. An increasing investment of energy in the defense of the ego eventually reaches a point of diminishing returns. The process detracts from the energy needed for other ego functions such as integration and reality mastering, thus a very high *Es* score would reflect weakness instead of strength. An empirical suggestion for the point of diminishing returns for clinic and hospitalized patients comes from the Barron (1953) and Wirt (1955) prognosis studies. Groups above or below a raw *Es* score of about 49 showed less improvement than those near this score.

The significantly higher *Es* scores for severe delinquents relates to Barron's (1956) finding of a cluster of *Q* sort items descriptive of aggressiveness, power-orientation, and disregard of the rights of others, which correlated positively with *Es* scores on his Air Force officer sample. If added to this line of reasoning are the high mean *Es* for the salesmanagers together with the knowledge of above average scores on *Pd* for them, the delinquents, and the officers, another interpretation of *Es* is suggested. The scale may be a sign of psychic energy per se, whether it be thought of as coming from the ego or the id. The inference as to source in any particular instance would be based on the case history materials and the rest of the MMPI profile. Without additional information, psychopaths might be false positives in selection of good bets for psychotherapy.

Using *Es* scores in profile analysis to supplement information provided by the other scales is suggested by some of the data that now exist. A combination of high *K* with high *Es* in nonhospitalized subjects implies defensiveness with the intent to appear socially desirable. In psychiatric patients the same pattern suggests ego syntonic behavior and a refusal to undertake psychotherapy or, if undertaken, lack of cooperation in the process. A high *Es* combined with a high *Pd* suggests freedom from intrapsychic conflict; the aggression and hostility in the personality are ego syntonic. A low *Es* combined with a high *Pd* suggests that the hostility is not ego syntonic and that the subject may be an acting out neurotic. All of these clinical hunches have yet to be verified but are suggestive of the uses to which the *Es* scale might be put.

In conclusion, the multiplicity of ego functions makes it difficult to assess ego strength by observing or knowing the state of a few of these functions. Adding to the problem is the fact that an excessive amount of energy devoted to one function could be pathological but would be manifested by a high score on the *Es* scale. It appears that energy not available to the ego may manifest itself on the *Es* scale. Consistent interpretations of what the scale does depend both upon the particular kind of subject and his test-taking attitude. Any valid technique for the assessment of ego strength must conceive of the ego as a complex system and then must combine the results of an intrasystemic analysis with intersystemic and interpersonal data.

Summary

More construct validation of the Ego-Strength scale has been described together with a rationale for the properties such a scale should ideally possess. The *Es* scale was administered to a variety of samples: superior adults, psychiatric adults, emotionally disturbed adolescents, normal adolescents, and severe delinquents. The scale broadly discriminated between psychiatric and nonpsychiatric groups. Suggestions for other interpretations of what the *Es* scale may be doing were made, as were suggestions for its use in profile analysis.

Received June 27, 1958.

REFERENCES

Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.

Barron, F. Ego-strength and the management of aggression. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956. Pp. 579–585.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

Freud, Anna. *The ego and the mechanisms of defense*. New York: Int. Universities Press, 1946.

Freud, S. The defence neuro-psychoses. In *Collected papers*. Vol. 1. London: Hogarth, 1924.

Freud, S. *The ego and the id*. London: Hogarth, 1927.

Hartmann, H. Comments on the psychoanalytic theory of the ego. In R. S. Eissler et al. (Eds.), *The psychoanalytic study of the child*. Vol. V. New York: Int. Universities Press, 1950. Pp. 74–97.

Hathaway, S. R., & Monachesi, E. D. (Eds.) *Analyzing and predicting juvenile delinquency with the MMPI*. Minneapolis: Univer. Minnesota Press, 1953.

Hathaway, S. R., & Briggs, P. F. Some normative data on new MMPI scales. *J. clin. Psychol.*, 1957, **13**, 364–369.

Nunberg, H. Practice and theory of psychoanalysis. *Nerv. ment. Dis. Monogr.*, 1948, No. 74.

Tamkin, A. S. An evaluation of the construct validity of Barron's ego-strength scale. *J. clin. Psychol.*, 1957, **13**, 156–158.

Tamkin, A. S., & Klett, C. J. Barron's Ego-Strength Scale: A replication of an evaluation of its construct validity. *J. consult. Psychol.*, 1957, **21**, 412.

Welsh, G. S. An extension of Hathaway's MMPI profile coding system. *J. consult. Psychol.*, 1948, **12**, 343–344.

Wirt, R. D. Further validation of the ego-strength scale. *J. consult. Psychol.*, 1955, **19**, 444.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

VALIDATION OF A NEED SCALING TECHNIQUE FOR THE ADJECTIVE CHECK LIST

ALFRED B. HEILBRUN, JR.
State University of Iowa

The Adjective Check List (ACL) is a test devised by Gough (1955) for personality research. It contains 300 adjectives which can be selectively checked as being characteristic of an *S*'s behavior. Gough suggests three advantages of this type of measure—a wide scope of behavior can be evaluated, the familiar words provide a meaningful task for the rater, and the presence-or-absence checking response assures analytic ease. However, one disadvantage of the check list as it is typically used is that it provides information which is conceptually circumscribed. For example, research may determine that two groups of *S*s are distinguishable by the differential frequency of several adjectives checked as characteristic. Although of interest, such findings offer limited basis for behavior prediction or theory construction.

One method of increasing the conceptual utility of the ACL would be to score responses by adjective clusters measuring new and broader behavior dimensions. LaForge and Suczek (cited in Gough, 1950) did this by using a sixteen-fold classification of interpersonal behaviors, while researchers at the Institute for Personality Assessment and Research (Gough, 1950, p. 38) developed clusters intended to measure self-acceptance and self-criticality.

Another approach was taken by Heilbrun (1958) who developed 15 need scales for the ACL, utilizing needs initially described by Murray (1938) and later incorporated into the Personal Preference Schedule (PPS) by Edwards (1954). It was found that the rank order correlation between relative need levels as evaluated by the need scales and the PPS was .60, significant at the .05 level. With personal desirability held constant, the first-order partial correlation dropped to .35, which did not differ significantly from a zero correlation. Since there has been a limited amount of published validational data for the PPS, the rather low order correlation did not necessarily deny the usefulness of the ACL need scales. However, it did point out the necessity of relating need scores with external (nontest) criteria in way of evaluating their current utility for personality research. The present study assessed the validity of five of the 15 need scales by relating test performances to such external criteria.

Procedure

Subjects

The *S*s for this study were 99 members of an undergraduate college course in psychology. This number included 56 females and 43 males.

Task

The present scales for the ACL were first rationally derived by having 20 advanced graduate students judge which adjectives, if checked, would indicate a high level of the following needs: achievement, deference, order, exhibition, autonomy, affiliation, intraception, succorance, dominance, abasement, nurturance, change, endurance, heterosexuality, and aggression. Agreement among 9 of the 20 judges was required for inclusion of an adjective on a need scale. The number of adjectives on these scales ranged from 19 (succorance) through 33 (aggression). The resultant adjective clusters were those used in a previous study (Heilbrun, 1958). These scales were later revised by having 19 advanced graduate students judge which adjectives, if checked, would contraindicate a high level of a given need. Agreement among 9 of the 19 judges was required for inclusion.

The number of contraindicative adjectives for each scale ranged from 7 (intraception) to 36 (aggression).[1]

A need score on the ACL was obtained by subtracting the number of contraindicative adjectives checked by *S* from the number of adjectives indicative of a high need level which were endorsed as self-descriptive.

Procedure

All members of an undergraduate psychology course were given the ACL under standard instructions on the first day of class. Approximately three months later factual information was obtained by questionnaire from the *S*s and from university records. These data were selected as validating criteria for five need scales—achievement, exhibition, affiliation, nurturance, and abasement. The behaviors which will be presented as descriptive of these needs are taken from Edwards (1954) and were used by the judges in deriving the need scales.

Results

Achievement

The following behaviors characterize the achievement need:

> To do one's best, to be successful, to accomplish tasks requiring skill and effort, to be a recognized authority, to accomplish something of great significance, to do a difficult job well, to solve difficult problems and puzzles, to be able to do things better than others, to write a great novel or play.

The criterion for achievement need was college grade point average (GPA) with an *S*'s estimated intelligence held constant. The intellectual estimate was made from a vocabulary test included in the entrance battery administered to all entering freshmen at the State University of Iowa. The distributions of available GPAs and vocabulary percentile scores were made for the experimental *S*s and each distribution was divided into quartiles. A high achiever (HA) was defined as an *S* whose GPA rank was at least two quartiles above his vocabulary rank. Those *S*s who appeared in the third quartile of the vocabulary distribution and the fourth quartile of the GPA distribution were not considered nor were those *S*s considered who fell in the fourth quartiles, respectively, of these distributions. The rationale for elimination was that in neither case was it possible for an *S* to qualify as an HA because of the high vocabulary standing.

Table 1

Comparison of Means for Validation of Five ACL Need Scales

Need	*N*	Mean	*SD*	*t*	*df*	*p*
Achievement						
High	9	12.89	5.93	2.12	58	.04
Low	51	8.51	5.82			
Exhibition						
High	24	14.13	4.84	2.13	44	.04
Low	22	10.95	4.89			
Affiliation						
High	25	15.04	7.36	2.17	47	.03
Low	24	10.17	7.79			
Nurturance						
High	32	5.22	1.78	2.29	66	.03
Low	36	4.05	2.34			
Abasement						
High	26	−1.04	4.92	2.69	48	.01
Low	24	2.67	4.71			

The performances of HAs on the Achievement scale were compared with the remaining *S*s, and the results are presented in Table 1. It can be seen that the high achievement group scored significantly higher on the Achievement scale than did the remainder of the experimental *S*s.

Exhibition

Edwards describes exhibition as follows:

> To say witty sayings, to tell amusing jokes and stories, to talk about personal adventures and experiences, to have others notice and comment upon one's appearance, to say things just to see what effect it will have on others, to talk about personal achievements, to be the center of attention, to use words that others do not know the meaning of, to ask questions others cannot answer.

It is obvious from these descriptive behaviors that a person with a high exhibition need would seek social interactions with others, since their responses would be necessary for need satisfaction. It seems reasonable to assume that such a person would be more group oriented than would a person without high need exhibition. Therefore, the validating criterion for the Exhibition scale was the number of group activities (social, academic, athletic, church, etc.) of which *S* had been a voluntary member from the time he had entered high school until the present. The error introduced by age differences between *S*s, which would provide a greater time span since entering high school, was minimal because of the homogeneous age make-up of the experimental group ($M = 21.08$; $SD = 3.09$). Also an appreciable relationship between age and Exhibition scale scores would be necessary before any systematic error would be introduced, and a performance comparison of the 22 youngest and oldest *S*s on this scale did not support this contention ($t = 0.11$; $p = .92$).

---

[1] These need keys may be obtained without charge from Alfred B. Heilbrun, Jr., Department of Psychology, State University of Iowa, Iowa City, Iowa.

The number of groups joined by those *S*s who fell in roughly the highest and lowest quartile of scores on the Exhibition scale was compared. The comparison, shown in Table 1, shows that high scorers on the scale were members of a reliably greater number of groups than were low scorers.

Affiliation

Behaviors descriptive of a high affiliation need are:

> To be loyal to friends, to participate in friendly groups, to do things for friends, to form new friendships, to make as many friends as possible, to share things with friends, to do things with friends rather than alone, to form strong attachments, to write letters to friends.

The criterion which was selected for the Affiliation scale was the number of friends which an *S* had. Experimental *S*s were asked to list the names of people they had met within the previous five years whom they would consider "good friends." This definition of a "good friend" was provided: "A person whose company you seek and enjoy and with whom you would be willing to discuss important *personal* experiences." Relatives were not included.

The mean number of people considered to be "good friends" by *S*s whose scores fell within roughly the highest and lowest quartiles on the Affiliation scale is given in Table 1. These data demonstrate a significantly greater number of people considered to be "good friends" by *S*s scoring highest on the Affiliation scale than by *S*s scoring lowest.

Nurturance

Nurturant behaviors are described in this manner:

> To help friends when they are in trouble, to assist others less fortunate, to treat others with kindness and sympathy, to forgive others, to do small favors for others, to be generous with others, to sympathize with others who are hurt or sick, to show a great deal of affection toward others, to have others confide in one about personal problems.

Validation of the Nurturance scale was attempted by relating it to the number of charitable, medical research, rehabilitation, church, or educational activities to which *S* had contributed time, money, or personal effects within the previous two years. The mean frequencies of contributions made by *S*s falling approximately in the upper third and lower third of scores on the Nurturance scale are shown in Table 1. It can be seen that *S*s scoring high on the Nurturance scale made reliably more contributions than did the lower scoring *S*s.

Abasement

Characteristic behaviors associated with a need for abasement are as follows:

> To feel guilty when one does something wrong, to accept blame when things do not go right, to feel that personal pain and misery do more good than harm, to feel the need for punishment for wrong doing, to feel better when giving in and avoiding a fight than when having one's way, to feel depressed by inability to handle situations, to feel timid in the presence of superiors, to feel inferior to others in most respects.

The validating procedure involved the prediction of expected course grade by the experimental *S*s at the first class meeting. The predicted grade for each *S* was compared with his then current cumulative grade point average, these differences providing the criterion measure. The specific prediction made was that *S*s who estimated course grades markedly in excess of their current average would show lower scores on the Abasement scale.

The distribution of discrepancy scores was

Table 2

Significance of Differences in "Irrelevant" Criterion Performance of Approximately the Highest and Lowest Quartiles of *Ss* on the Various Need Scales

"Irrelevant" Criterion	Achievement Scale	Exhibition Scale	Affiliation Scale	Nurturance Scale	Abasement Scale
	t	*t*	*t*	*t*	*t*
Achievement		1.34	.80	.74	.26
Exhibition	.66		2.63*	.78	.71
Affiliation	.59	1.90		1.88	2.27**
Nurturance	1.14	.43	1.82		1.01
Abasement	.18	.67	.71	.73	

* Significant at $p < .01$ level.
** Significant at $p < .03$ level.

separated at roughly the highest and lowest 40%, high in this case referring to predicted course grades markedly above grade point average. The mean Abasement scale scores for these two groups are found in Table 1. These data show, as predicted, that *Ss* with high discrepancy scores were reliably lower on the Abasement scale than *Ss* with low discrepancy scores.

Adequacy of the Criteria

Despite the consistent finding of significant relationships between the five need scales and the selected criteria, one further analysis was necessary before the results could be accepted as validating this mode of personality assessment. It should be demonstrated that the selected criteria are relevant. One method of evaluating relevancy would be to determine the relationships between each criterion and the four scales which were *not* validated against it. These relationships should be less clear than were those previously reported between need scales and "relevant" criteria. The check was made by comparing the "irrelevant" criterion performance of roughly the highest and lowest quartiles of *Ss* on each scale. An exception to this procedure was made for the Achievement scale criterion which was not a continuous variable. In this case, the performances of the criterion groups (nine high achievers, 51 remaining *Ss*) on the "irrelevant" scales were compared. The results of this analysis are presented in Table 2. It can be seen that only two of the 20 relationships reached the .05 level of significance (one would be expected by chance). This would seem to support the contention that the variables selected for this study are relevant validating criteria.[2]

Discussion

The finding that performance on each of the five need scales for the ACL was significantly related to its criterion variable gives rather clear support to this technique of personality assessment. It should be stressed that the ability of a scale to reliably discriminate between extremes on a criterion dimension does not necessarily mean that such an instrument would be useful for individual prediction. However, since the ACL and the need scales which are currently being investigated are intended as research tools, individual prediction is not a crucial factor. The status of current psychological research is such that group predictions or group discriminations are the rule rather than the exception. The question of whether this technique of adjective endorsement has potential clinical utility would certainly seem to be premature, if not presumptuous, at this stage of development. The research utility of an easily administered, short, objectively scored measure of motivational status is evidenced by the expanding literature on the Manifest Anxiety scale developed by Taylor (1956). The current goal in developing the need scales is to expand the possibilities for group motivational research.

---

[2] A sixth need scale, dominance, was initially included in this validational study but was eliminated following the analysis of relationships of the selected criteria with "irrelevant" scales. Although the dominance scale demonstrated the predicted relationship with number of elected group offices held, high achievement, affiliation, and nurturance *Ss* also reported significantly ($p < .05$) more elected offices held than did *Ss* low on these scales. Accordingly, the criterion was not deemed an adequate validating dimension for the dominance scale.

Since four of the five validating criteria were obtained by verbal report of the experimental *S*s, the question of dissimulation could be raised. There are two factors which would seem to curtail the influence of this source of error. The first is that *E* was also the classroom instructor, and it was felt that a rather good rapport existed at the time the autobiographical material was obtained. The second factor is that *S*s were required to make their responses explicit by naming groups, friends, and recipients of donations. Accordingly, distortion under these circumstances would have to approach outright prevarication. In light of the rapport, promised confidentiality of the material, and a frank exhortation for honesty by *E*, deliberate distortion should not have been a major source of error.

Dissimulation on the ACL, when it is administered as a self-descriptive task, still remains a problem. It has been demonstrated (Heilbrun, 1958) that frequency of endorsement for adjective need clusters is highly related to judged social desirability of the adjectives characteristic of these needs. The extent to which this source of test performance variance contributes to predictive error is not clear, since it seems defensible to assume that some human social behavior is also highly related to a person's judged desirability of the behavior in question. Thus, merely because test performance is determined to some extent by desirability of test alternatives does not preclude behavioral prediction. However, the extent of the correlation between frequency of adjective endorsement and judged social desirability of the adjectives ($\rho = .98$) suggests that error is involved. Further research is necessary to evaluate the magnitude of testing error introduced by desirability factors and to investigate corrective procedures on the ACL.

Summary

This study related performance of 99 college *S*s upon five rationally derived need scales for the Adjective Check List to appropriate validating criteria in way of evaluating the usefulness of this mode of objective personality appraisal.

High achieving *S*s, as defined by relatively high grade point averages with estimated intelligence held constant, scored significantly higher on the Achievement scale than did *S*s who did not show this characteristic.

High scorers on the Exhibition scale had entered into significantly more group activities than had *S*s scoring low on this scale.

*S*s with high scores on the Affiliation scale had met a reliably greater number of people within the previous five years whom they described as "good friends" than had *S*s with low scale scores.

High performers on the Nurturance scale had made a reliably greater number of contributions of time, money, or personal effects than low performers.

*S*s who, when asked to predict their course grades, made predictions markedly in excess of their current grade point averages scored significantly lower on the Abasement scale than *S*s who showed smaller discrepancies.

Adequacy of the validating criteria was indicated by the failure to find clear relationships between these criteria and those scales for which they were "irrelevant" in this study.

The findings were seen as indicating the potential usefulness of this technique of objective need measurement in personality research.

Received June 27, 1958.

REFERENCES

EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

GOUGH, H. G. *Predicting success in graduate training: A progress report.* Berkeley: Univer. California Inst. Pers. Assess. & Res., 1950.

GOUGH, H. G. *Reference handbook for the Gough Adjective Check-List.* Berkeley: Univer. California Inst. Pers. Assess. & Res., 1955. (Mimeo.)

HEILBRUN, A. B. Relationships between the Adjective Check-List, Personal Preference Schedule, and desirability factors under varying defensiveness conditions. *J. clin. Psychol.*, 1958, **14**, 283–287.

MURRAY, H. A. Explorations in personality: A clinical and experimental study of fifty men of college age. New York: Oxford Univer. Press, 1938.

TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

THE SCREENING VALUE OF THE CORNELL MEDICAL INDEX

M. POWELL LAWTON [1]
Norristown State Hospital, Norristown, Pennsylvania

The Cornell Medical Index Health Questionnaire (CMI) has seen wide use as a device for screening possible psychiatric patients. In the author's own experience it has proved particularly useful because, among all self-administering questionnaires, it seems least threatening to the average person. The nonthreatening quality, which undoubtedly is due to the medical orientation of most of the questions, is especially valuable when one must select emotionally disturbed persons from a population which includes many average people with a layman's prejudices and fears regarding emotional illness. Particularly efficient use may be made of such a short, nonthreatening test when a group is to be tested for research purposes and there is not time for individual interview and rating of patients. Further information is needed, however, as to whether the CMI actually discriminates the reasonably normal person from the clinically maladjusted person, and whether it discriminates among degrees of maladjustment.

The authors of the test have provided some confirmation of the former question (Brodman, Erdmann, Lorge, Deutschberger, & Wolff, 1954; Brodman, Erdmann, Lorge, Gershenson, & Wolff: 1952a, 1952b) by distinguishing psychiatric general hospital patients from other hospital patients, and psychiatric army "rejects" from psychiatric "accepts." The second question, which has not been investigated, is particularly relevant if the test is to be used as a criterion for extent of psychopathology in research populations, as suggested by the authors (Brodman, Erdmann, & Wolff, 1955, p. 3).

The present study attempted to determine (*a*) whether the CMI actually discriminates degrees of maladjustment, (*b*) what are the best indices for such discrimination, and (*c*) whether chronic illness is a variable that tends to weight scores in the direction of psychopathology.

Procedure

A rough order of "likelihood of nervous symptoms" was arrived at in the following manner. CMI blanks were given to 116 consecutive male veteran admissions to a general hospital (excluding those who were critically ill or uncooperative), and to 34 consecutive nonpsychotic admissions to the psychiatric service. The treatment files on each patient for present, as well as past, admissions to the hospital were consulted, and after careful reading of all available data the patient was placed in one of the following four categories:

I. Patients whose records contained no intimation of any psychological disturbance or the presence of any illness considered "psychosomatic" ($N = 43$).

II. Patients whose records contained either some notation by a physician of nervousness or tension [2] ($N = 17$) or an official diagnosis of an illness mentioned as showing prominent emotional concomitants by Weiss and English (1949) ($N = 21$).[3] Since such a spontaneous evaluation by a physician might refer to a "nervous" trait either in the normal or the pathological range, and since a "psychosomatic" diagnosis implies greater likelihood of

---

[1] This study was initiated while the author was at the Veterans Administration Hospital, Providence, Rhode Island.

[2] Typical physicians' comments were "marked emotional lability," "feels tense and jittery," "a rather nervous individual."

[3] This psychosomatic group consisted of 15 cases of peptic ulcer, 9 essential hypertension, 3 diabetes mellitus, 3 bronchial asthma, and one migraine headache.

Table 1

Age Composition of the Experimental Groups

Group	N	Mean	Sigma
I	43	39.5	15.71
II	48	43.5	14.93
III	25	42.6	11.52
IV	34	40.2	11.94
All groups	150	41.5	14.07

Table 2

Cumulative Percentages of Total Score by Groups

Score	Group I	Group II	Group III	Group IV
10	74	85	92	97
20	42	62	72	85
25	40	47	52	82
30	26	40	48	76
35	19	34	40	71
40	19	17	40	71
50	9	11	28	62
60	5	6	8	35
70	2	2	4	18

emotional symptomatology than absence of such medical diagnosis, it was believed that a good screening device should detect more maladjusted individuals in Group II than in Group I.

III. Patients who were at that time hospitalized on the medical and surgical services and whose records indicated an official diagnosis of a psychiatric condition, either for the previous hospitalization or on previous occasions ($N = 25$). It was reasoned that the patient whose disturbance was marked enough to warrant an official psychiatric diagnosis was more likely to show a wide range of symptoms than patients in Group II.

IV. Patients at the time hospitalized on the psychiatric service ($N = 34$). They should logically show a wider range of disturbances than those in Group III, whose psychiatric disability was probably not causing their hospitalization at the time of taking the test. Whenever a patient's record showed descriptions applying both to Group II and either Group III or Group IV, he was placed in the group of greater maladjustment. Such an ordering of degree of maladjustment, while quite rough, seems logically defensible and has the advantage of being unbiased and relatively objective.

Table 1 lists the composition of the groups. An analysis of variance and Bartlett's test indicate that the age means and variances do not differ significantly among groups. Inasmuch as education has been shown to be unrelated to CMI scores (Brodman, Erdmann, Lorge, & Wolff, 1953), it was not considered in this study.

Results

The authors suggest (Brodman et al., 1955) that psychiatric illness be suspected if the patient's CMI protocol is characterized by one or more of the following signs: (*a*) Total Score (number of Yes responses) 30 or more; (*b*) Page 4 Score 3 or more ("Moods and feelings" items on page 4 of the questionnaire); (*c*) Sections I and J Score 3 or more (questions dealing with fatigue and frequency of illness); (*d*) Omit, etc., 4 or more (Questions Omitted, answered both Yes and No, or with changes or comments added). In addition, the discriminating power of the following indices of maladjustment was evaluated by classifying as pathological any patient who showed (*e*) Areas Score—7 areas or more among the "body complaints" with more than one Yes response (questionnaire pages 1–3); (*f*) Page 1–3 Score 28 or more (body complaint questions).

(*a*) *Total Score*. The cumulative percentage distribution of Total Score is indicated in Table 2 for the four patient groups. For every

Table 3

Cumulative Percentages of Page 4 Scores by Group

Score	Group I	Group II	Group III	Group IV
0	60	67	84	94
1	51	60	76	88
2	35	50	60	79
3	26	48	54	79
4	21	42	44	79
5	21	37	44	79
6	19	31	40	79
7	19	31	40	79
8	16	27	24	76
9	16	23	16	71

Table 4

Median Scores on C.M.I. Measures and Percentages of Patients Pathologically Diagnosed by Each Measure

	Measure											
	Total Score		Page 4		Sect. I & J		Omit, etc.		Areas		Page 1–3	
Group	Median	%	Median	%	Median	%	Median	%	Median	%	Median	%
I	18	26	2	35	0	30	0	16	4	33	15	23
II	23	40	2	50	1	35	1	23	5	45	18	34
III	26	48	4	60	3	54	1	24	6	64	22	40
IV	53	76	17.5	79	7	71	2	35	8	76	34	59
Chi square	20.4		15.8		11.7		3.84		16.9		10.5	
p	<.001		<.01		<.01		>.05		<.001		<.02	

cutting point but one (40 or more), until the frequencies become too small for significance, this index orders the scores in the predicted manner: the percentage of Yes responses becomes greater as the groups increase in likelihood of emotional illness.

(*b*) *Page 4 Score*. Table 3 lists the cumulative percentages of Yes responses on Questionnaire Page 4. Until a cutting point of eight is reached, this section of the questionnaire succeeds in ordering the four groups as predicted.

(*c*) *Sections I and J Score*. These results are tabulated in Column 3 of Table 4, which is a summary table of all the specific scoring criteria that were evaluated. This index appears to discriminate well among the groups.

(*d*) *Omit, etc*. Column 4 of Table 4 indicates the relative occurrence of this sign. Differences among groups, though in the predicted order, appear to be statistically random. However, since the over-all frequency of occurrence was low, it was decided to dichotomize the distribution at the median, which was one. Using this cutting point, Group I showed 27% at or above the median, Group II showed 60% above, Group III 52%, and Group IV 79%. These differences were significant at the .01 level, but the predicted direction of differences was reversed for Groups II and III.

(*e*) *Areas Score*. Column 5 of Table 4 shows that this index reliably distinguishes the groups.

(*f*) *Page 1–3 Score*. Column 6 of Table 4 indicates that this index, while still giving significant discrimination among groups, was less efficient than either the Total Score or the Page 4 Score.

Chronic illness was defined in a rough, arbitrary fashion as being either (*a*) episodic illness severe enough to have caused at least two hospitalizations previous to the present one or (*b*) the present illness having currently or at some previous time lasted for a period of one year or more, with some vocational disability. By these criteria, 23 patients out of the 116 comprising Groups I, II, and III were classified as having chronic illness. A chi square test indicated that the proportion of Group I patients in the chronic illness group did not differ significantly from the proportion of Group I patients in the entire group. Therefore, there was no difference between the chronically ill and the nonchronically ill in the presence of overt psychopathology. Each patient having a chronic illness was matched with a patient without chronic illness of the same group and age. In total score, 52.1% of chronic patients scored above the median, while 47.9% of the nonchronic patients scored above the median ($p > .05$). No difference was found between chronic and nonchronic patients in Page 4 Score ($p >$ .05), also using the median test (Moses 1952).

Discussion

The CMI is apparently able to distinguish between degrees of emotional illness. It seems reasonable to think that had a more exact criterion of maladjustment been used, the discrimination among groups would have been even sharper.

Inspection of Table 4 reveals that for the present sample, best discrimination among all groups was provided by the total score and the Page 4 Score. The frequency distribution of Table 2 indicates that the use of total score of 35, 40, or 50 as cutting points would heighten the discrimination between known psychiatric and presumably nondisturbed patients. A cutting score of 50 identified 60% of this sample of known male psychiatric patients, while in other studies (Brodman et al., 1952a; Arnhoff, Strough, & Seymour, 1956) this score identified only about 45% and 36%, respectively. Since both of these earlier studies used outpatients, the higher scores of the present group could conceivably reflect more severe illness in the hospitalized psychiatric population.

On the other hand, if best discrimination among patients admitted for medical and surgical conditions (Groups I, II, and III) is desired, the use of a cutting score of 20 is indicated by Table 2. The cutting score of 30 originally suggested by the test's authors (Brodman et al., 1955) would seem to give the best over-all discrimination. Likewise, from Table 3 it is seen that while cutting scores of 6, 7, or 8 on mood and feeling items give best separation of the extreme groups, the authors' suggested score of 3 seems to be preferable when attempting to pick out emotionally disturbed medical and surgical patients.

These results reinforce the suggestion that interpretation of screening devices be varied according to the population tested and the purpose for which they are used (Meehl & Rosen, 1955). For instance, the present results, when compared with other studies (Arnhoff et al., 1956; Brodman et al.: 1954; 1952a; 1952b), suggest tentatively that V.A. medical and surgical patients, psychiatric outpatients, and psychiatric inpatients may tend to receive higher scores than nonveteran populations. Though much larger samples would be needed to establish the existence of such broad population differences, the present study suggests that different cutting scores may be better for V.A. populations.

In view of the low frequency of occurrence of the Omit, etc., sign, and a failure to validate them when each was considered separately (Arnhoff et al., 1956), the present study indicates that lumping them all together and considering a single occurrence as possibly noteworthy may be of value. Probably this sign is most efficient as a gross indicator of psychopathology, inasmuch as the intermediate groups were not ranked in the predicted order.

Lack of information on the base rate of emotional disturbance in V.A. general hospitals makes interpretation of the predictive value of the CMI difficult. However, it is probable that a high proportion of admissions to general hospitals have some degree of observable psychopathology, judging by such data as Zwerling, Titchener, Gottschalk, Levin, Culbertson, Cohen, and Silver (1955). Even if it is assumed that all patients in Group I are, as the criterion indicates, psychiatric "normals," a maximum false positive rate of 20% was obtained (Table 4), while the false negative rate was 24% among Group IV. Under these conditions, use of the CMI would improve upon base rates if the proportion of maladjusted patients admitted were as low as 1:4 (Meehl & Rosen, 1955). The ability to pick out a fair proportion of probably maladjusted individuals from a large group makes the CMI valuable in one of its prime functions, that of serving as a criterion for the inclusion of subjects in "normal" research populations.

Finally, a preliminary determination of how and to what extent chronic physical illness is confounded with emotional illness in the elevation of CMI scores was made. In this group, when both age and degree of overt maladjustment were held constant, chronic illness in itself did not seem to inflate CMI scores. Should this conclusion be confirmed in future work it would add strength to the assertion that high CMI scores are mainly a function of psychopathology and that use of the CMI as a psychiatric screening device in a general medical setting does not risk inordinate numbers of false positives.

Summary

The value of the Cornell Medical Index as a psychiatric screening device in a general medical hospital was investigated. An ordinal

scale of presumed degree of emotional illness was obtained by classifying 116 consecutive medical and surgical admissions as (I) without evidence in their records of psychiatric disorder, (II) presence of clinically noted tension or psychosomatic disorder, (III) presence of clinically diagnosed psychiatric disorder. A further group of 34 consecutive psychiatric admissions constituted (IV) the "most probably ill" group. Various CMI scores as suggested by the authors were studied, and it was found that these scores consistently ordered the four groups in terms of measured psychopathology, as predicted. Most of the authors' suggested cutting points were confirmed, but there was some evidence that different cutting points would be valuable for different purposes. In a small sample of 23 cases, it was found that presence of chronic illness did not by itself increase the likelihood of psychiatric diagnosis by the CMI. It was concluded that the CMI is valuable both as a screening device for choosing groups and as an individual test.

Received July 7, 1958.

REFERENCES

ARNHOFF, F. N., STROUGH, L. C., & SEYMOUR, R. B. The Cornell Medical Index in a psychiatric outpatient clinic. *J. clin. Psychol.*, 1956, **12**, 263–268.

BRODMAN, K., ERDMANN, A. J., LORGE, I., DEUTSCHBERGER, J., & WOLFF, H. G. The Cornell Medical Index-Health Questionnaire VII. The prediction of psychosomatic and psychiatric disabilities in army training. *Amer. J. Psychiat.*, 1954, **111**, 37–40.

BRODMAN, K., ERDMANN, A. J., LORGE, I., GERSHENSON, C., & WOLFF, H. G. The Cornell Medical Index-Health Questionnaire III. The evaluation of emotional disturbances. *J. clin. Psychol.*, 1952a, **8**, 119–124.

BRODMAN, K., ERDMANN, A. J., LORGE, I., GERSHENSON, C., & WOLFF, H. G. The Cornell Medical Index-Health Questionnaire IV. The recognition of emotional disturbance in a general hospital. *J. clin. Psychol.*, 1952b, **8**, 289–293.

BRODMAN, K., EREMANN, A. J., LORGE, I., & WOLFF, H. G. The Cornell Medical Index-Health Questionnaire VI. The relation of patients' complaints to age, sex, race and education. *J. Geront.*, 1953, **8**, 339–342.

BRODMAN, K., ERDMANN, A. J., & WOLFF, H. G. *Cornell Medical Index Manual.* New York: Cornell Univer. Med. Coll., 1955.

MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, **52**, 194–216.

MOSES, L. E. Nonparametric statistics for psychological research. *Psychol. Bull.*, 1952, **49**, 122–143.

WEISS, E., & ENGLISH, O. S. *Psychosomatic medicine.* Philadelphia: Saunders, 1949.

ZWERLING, I., TITCHENER, J., GOTTSCHALK, L., LEVIN, M., CULBERTSON, W., COHEN, S., & SILVER, H. Personality disorder and the relationships of emotion to surgical illness in 200 surgical patients. *Amer. J. Psychiat.*, 1955, **112**, 270–277.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

SYMBOLISM VALIDITY AND LEARNING WITHOUT AWARENESS

CAROLINE TAYLOR MACBRAYER
Davidson College

In a recent publication, Levy (1954) reported evidence against the existence of a universal sexual symbolism and warned against the current practice of assuming, in blind analysis of projective materials, that universal sexual symbolism is a valid concept. Levy's statement (1954, p. 45) that his results should not be generalized to any great extent beyond his experimental population of fifth-grade children, plus continued and renewed emphasis on sexual symbolism in current research (Clark, 1955; Hall, 1953), were the chief stimuli for the present investigation. The two purposes of the present experiment are to determine whether Levy's (1954) negative conclusion concerning the existence of a universal sexual symbolism of the Freudian or classical type holds true for college students, and to determine whether learning of the meaning of symbols in relation to sex, without awareness of what is being learned, can be experimentally produced. This latter aim was suggested by current studies on learning without awareness of what is being learned (Cohen, Kalish, Thurston, & Cohen, 1954; Greenspoon, 1955; Philbrick & Postman, 1955; Postman & Jarrett, 1952), especially the experiment by Philbrick and Postman (1955), which showed that a relatively simple task facilitates learning without awareness.

Essentially, the present experiment is an investigation of the influence of unconsciousness, as defined in at least two of the 16 ways discussed by Miller (1942, pp. 21–44), on the responses of subjects (*S*s) to sexual symbols. A strictly psychoanalytic concept of unconsciousness (Miller, 1942, pp. 42–43) was used by Levy (1954), who found that unconscious classical or Freudian sexual symbolism did not produce significantly greater learning by the paired-associates technique in one group of fifth-grade children, for whom boys' first names were paired with abstract "male" figures (long, pointed) and girls' first names were paired with abstract "female" figures (round, containing), as compared with another group of fifth-grade children, for whom boys' names were paired with "female" figures and girls' names with "male" figures. Levy (1954) also found that these same children did not pair the five girls' names with the five female figures or five boys' names with the five male figures at a significantly better than chance level.

The present study is designed in a similar, but somewhat more extended and rigorous, manner to investigate this same psychoanalytic concept of unconscious sexual symbolism. It also attempts to investigate unconscious learning during the experiment; unconscious learning being defined here as learning without awareness of what is being learned (Miller, 1942, pp. 43–44), or learning without being able to communicate or verbalize what has been learned (Miller, 1942, pp. 38–41 and p. 292). Hypotheses to be tested by the experiment are as follows:

1. Classical or Freudian unconscious sexual symbolism is valid for male college students.
2. Learning of the characteristic or characteristics of maleness or femaleness of abstract figures can occur without awareness on the part of the *S* of what is being learned.

In addition to testing these two hypotheses, the present experiment aimed to determine, within the current experimental procedure, whether unconscious classical sexual symbolism, or learning during the experiment without awareness of what is being learned, is a stronger determiner of responses to abstract sexually symbolic figures.

Method

*S*s were 80 male college students in a men's college, none of whom had taken a college course in psychology. They were divided into four experimental groups, each of which contained 20 *S*s. The 13 freshmen, 23 sophomores, 34 juniors, and 10 seniors were assigned at random to one of the four experimental groups, the number of *S*s from each class in each of the four groups being roughly equivalent.

The procedure was the same for all groups except that the nature and order of materials presented to each group differed, as indicated below. Materials consisted of the 10 abstract figures[1] used by Levy (1954), five of which were "male" (elongated or pointed) and five "female" (rounded or containing) and the five boys' and five girls' first names which Levy used, plus 10 different abstract figures designed by the author,[2] five male and five female, with ten different first names—five of boys and five of girls. Each *S* performed two similar tasks of attempting to memorize names paired with figures, one task involving Levy figures and names, and the other, MacBrayer figures and names. For each *S*, boys' names were paired with male figures and girls' names with female figures in one memory task (called "correct" condition in the present experiment), while in the other memory task, boys' names were paired with female figures and girls' names with male figures (called "reversed" condition in the present experiment). The nature of materials and order of presentation for the four experimental groups, designed to counterbalance for practice effects and any difference in difficulty between Levy and MacBrayer figures and names, were as follows:

Group	*First Memory Task*	*Second Memory Task*
1.	MacBrayer figures, correct	Levy figures, reversed
2.	MacBrayer figures, reversed	Levy figures, correct
3.	Levy figures, correct	MacBrayer figures, reversed
4.	Levy figures, reversed	MacBrayer figures, correct

Materials were presented in mimeographed form. One minute was allowed for memorizing the names of the 10 figures in the first task, followed by a three-minute recall period during which the same figures were presented in a different order and *S* was asked to write the names of all the figures he could recall, guessing for those which he could not recall. At the end of this three-minute recall period, the same procedure, using *S*'s second memory task, was repeated. The experiment was presented to *S*s as a study to determine how much ability to memorize improves during the four years in college. Each *S* was asked to put his class but not his name on his paper. Immediately after completion of the test, *S*s were asked to comment on anything they had noticed about the experiment, such as the nature of the figures or anything they thought the test might measure in addition to ability to memorize. Four *S*s were eliminated from the experiment because they verbalized the principle at least partially, and four *S*s were eliminated because they made no errors under either correct or reversed conditions; the data for these *S*s are not included in the results of the 80 *S*s reported in the present experiment.

Results

Results are presented in Table 1. There is clearly no significant difference in memory for names of figures under correct as compared to reversed conditions. In other words, memory for names of figures is not significantly facilitated by boys' names being paired with male figures and girls' names with female figures, rather than vice versa. This agrees with Levy's (1954) findings for fifth-grade children, which was interpreted by him as evidence against universal sexual symbolism. Levy assumed that if universal sexual symbolism exists, positive transfer effects in memorizing would result from correct pairing of figures and names and negative transfer effects from reversed pairing. While these data reveal no evidence for classical sexual symbolism validity, an analysis of the nature of the errors in the present experiment, under "forced guessing" conditions, reveals a different story.

Under all conditions, a significantly greater number of guessed (that is, incorrect) names were of the same sex as the name which had been paired with the figure in the material *S* had attempted to memorize. This is interpreted as evidence of learning without aware-

---

[1] Two figures were slightly modified.

[2] Available on request from the author.

Table 1

Errors in Recall of Names of Abstract Figures

(Ss: $N = 80$)

Condition	M No. or Proportion of Errors		t or z	p
Names correct (c) vs. names reversed (r)	$M_c = 2.75$	$M_r = 3.01$	$t = .72$	—
Same-sex errors (s) vs. opposite-sex errors (o)[a]				
All conditions	$M_s = 1.87$	$M_o = 1.01$	$t = 4.71$	$<.01$
Correct condition	$M_s = 1.95$	$M_o = .80$	$t = 4.55$	$<.01$
Reversed condition	$M_s = 1.79$	$M_o = 1.22$	$t = 2.30$	$<.05$
Proportions of opposite-sex errors under correct (c) vs. reversed (r) conditions	$p_c = .29$	$p_r = .41$	$z = 3.64$	$<.01$

[a] A same-sex error is one in which the guessed name of the figure is of the same sex as the name paired with the figure in the material S attempted to memorize; an opposite-sex error is one in which the guessed name of the figure is of the opposite sex from the name paired with the figure in the material S attempted to memorize.

ness of what has been learned, i.e., *Ss* learned that figures had the characteristic or characteristics of maleness or femaleness, as evidenced by the sex of their guessed names for the figures, without being able to verbalize the reason or principle on which their guesses were based. This finding supports the second hypothesis.

Further analysis of same-sex and opposite-sex errors reveals that under correct conditions, the number of same-sex errors (guessed name of figure same sex as in memory task) was significantly greater than opposite-sex errors, at better than the .01 level, whereas under reversed conditions, this difference is less marked, being significant at better than the .05 level, but below the .01 level of confidence. Translated into proportions of opposite-sex errors under correct as compared to reversed conditions (see Table 1), this difference is highly significant. This is interpreted as supporting the first hypothesis, that classical or Freudian unconscious sexual symbolism is valid for male college students. Apparently, when *Ss* cannot recall the name of a figure, there is no interference by classical unconscious symbolism with guessing names of figures which are the same sex as the figure name in the memory task, *if* the memory task paired girls' names with female figures and boys' names with male figures. If, however, the paired names in the memory task were reversed, unconscious sexual symbolism *does* interfere with guessing names of figures which are of the same sex as names paired with the figures in the memory task, and a significantly greater proportion of opposite-sex errors occurs under reversed conditions. These opposite-sex errors under reversed conditions are, of course, errors in which girls' names are guessed for female figures and boys' names for male figures, in accord with classical sexual symbolism, despite the fact that in the memory task the sex of the names given to the figures had been reversed.

Discussion

Evidence for the validity of unconscious sexual symbolism was found in the present experiment. Apparently, its operation is at a deeper level of unconsciousness than that investigated by Levy (1954). Its influence on responses to sexual symbols was found to be relatively weaker than the influence of unconscious learning, or learning without awareness, which took place during the experiment. Thus, when *Ss* have attempted to learn names for figures which are of the opposite sex than the classical sexual symbolism of the figure and are required to guess for unrecalled names of figures, their guessed names are significantly more often of the same sex as those they attempted to memorize. At the same time, opposite-sex errors intrude significantly more often under these conditions than when paired names in the memory task are of the same sex as the classical sexual symbolism of the figure.

That unconscious learning during the experiment was a stronger determiner of the sex of guessed names of abstract figures in the experiment than was classical unconscious sexual symbolism may be due to several factors. Learning during the experiment, although *Ss* were unaware of or unable to verbalize what had been learned, had the advantage of greater recency and probably of greater motivation (Miller, 1940), in the sense that *Ss* consciously wished to do well, which would be an indication to themselves of their own ability or intelligence. Furthermore, classical sexual symbolism has the disadvantage of repression, at least during ordinary waking conditions, whether repression is regarded in classical psychoanalytic terms (Miller, 1942, pp. 234–239) or simply as avoidance conditioning (Eriksen & Kuethe, 1956). While it is certainly impossible to say how deeply unconscious classical sexual symbolism is, it seems safe to assume that in the hierarchical continuum of imperceptible gradations between consciousness and unconsciousness (Miller, 1942, pp. 136–137), it lies further below the limen of awareness than the learning without awareness which occurred during the experiment, especially as only the latter reached awareness, or became communicable, for the four *S*s who were eliminated from the experiment.

This interpretation of the present learning data agrees *in principle* with studies of discrimination without awareness, such as that of Miller (1933). These studies nearly all show that ". . . the reliability of the subjects' judgments increases directly with the intensity of the stimuli. If a valid extrapolation can be drawn from this finding, it would be that accuracy of perception increases as the stimulation approaches a supraliminal level" (McConnell, Cutler, & McNeil, 1958, p. 231). Apparently, unconscious learning during the experiment more nearly approached a supraliminal level of awareness than did classical unconscious sexual symbolism, which accounts, at least in part, for the greater influence of the former on responses to abstract sexually symbolic figures in the present experiment. This suggests that responses to projective materials which are projections of unconscious needs and desires may be more dependent on recent learning, or learning which is near the *S*'s limen of awareness, than on more deeply unconscious or repressed processes. The present study does not, of course, furnish evidence that this is necessarily true.

Summary

An experiment was done to determine whether classical sexual symbolism is valid for male college students. No evidence was found that memory for names of abstract figures was better when boys' names were paired with "male" (elongated or pointed) abstract figures and girls' names with "female" (rounded or containing) abstract figures than when the names were paired with the figures in a reverse manner. However, evidence in support of the validity of classical sexual symbolism was found in the nature of the errors or guesses of names for figures which *S*s could not recall. Even stronger evidence was found of the influence of learning during the experiment without awareness of what was being learned. Possible reasons were discussed for the relatively greater influence of unconscious learning during the experiment than of unconscious classical sexual symbolism on responses during the experiment to abstract sexually symbolic figures.

Received July 7, 1958.

REFERENCES

COHEN, B. D., KALISH, H. I., THURSTON, J. R., & COHEN, E. Experimental manipulation of verbal behavior. *J. exp. Psychol.*, 1954, **47**, 106–110.

CLARK, R. A. The effects of sexual motivation on phantasy. In D. C. McClelland (Ed.), *Studies in motivation*. New York: Appleton-Century-Crofts, 1955. Pp. 44–57.

ERIKSEN, C. W., & KUETHE, J. L. Avoidance conditioning of verbal behavior without awareness: A paradigm of repression. *J. abnorm. soc. Psychol.*, 1956, **53**, 203–209.

GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 406–416.

HALL, C. S. A cognitive theory of dream symbols. *J. gen. Psychol.*, 1953, **48**, 169–186.

LEVY, L. H. Sexual symbolism: A validity study. *J. consult. Psychol.*, 1954, **18**, 43–46.

MCCONNELL, J. V., CUTLER, R. L., & MCNEIL, E. B. Subliminal stimulation: An overview. *Amer. Psychologist*, 1958, **13**, 229–242.

MILLER, J. G. Discrimination without awareness. *Amer. J. Psychol.*, 1933, **16**, 831–842.

MILLER, J. G. The role of motivation in learning without awareness. *Amer. J. Psychol.*, 1940, **53**, 229–239.

MILLER, J. G. *Unconsciousness*. New York: Wiley, 1942.

PHILBRICK, E. B., & POSTMAN, L. A further analysis of "learning without awareness." *Amer. J. Psychol.*, 1955, **68**, 417–424.

POSTMAN, L., & JARRETT, R. F. An experimental analysis of "learning without awareness." *Amer. J. Psychol.*, 1952, **65**, 244–255.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

THE SOCIAL DESIRABILITY OF TRAIT DESCRIPTIVE TERMS:

APPLICATIONS TO A SELF-CONCEPT INVENTORY[1]

EMORY L. COWEN AND PHOEBUS N. TONGAS[2]
University of Rochester

One consequence of recent research on the formal attributes of verbal response behavior is the growing conviction, by those concerned with problems of assessment, that subjects' (*Ss*') responses to personality inventories may be influenced by factors quite extraneous to item content. Illustratively, Edwards (1953) reports a correlation of .87 between the probability of endorsement of a personality descriptive statement and its social desirability (SD). Despite such empirical findings, the number of available personality inventories seems to proliferate rapidly. This, in part, may reflect a seeming simplicity in putting together a series of verbal propositions ostensibly tapping almost any type of psychological variable. Considerably more difficult is the task of demonstrating that the variable in question is actually being measured by the verbal responses given to the items purporting to measure it.

With the development of "modern phenomenology" (Rogers, 1951; Snygg & Combs, 1949) there has followed the inevitable rash of new instruments (Berger, 1952; Bills, Vance, & McLean, 1951; Brownfain, 1952; Fey, 1954; Phillips, 1951) presuming to assess variables of prime import within this frame of reference. Typical of this group of new measures is the Bills Index of Adjustment and Values (Bills, 1958). This test consists of a series of 49 trait descriptive adjectives, each of which is rated three times—for self concept, self-acceptance, and ideal self—on a series of five-point scales. The assumption is made, based on the underlying theory, that the higher the self concept (SC), self-acceptance (SA), and ideal self (IS), the better adjusted is the *S*. A fourth score, the maladjustment index, is derived from the summed discrepancy (D score), without respect to sign, between SC and IS. Since a discrepancy between SC and IS is taken as an index of conflict or emotionality with respect to the trait in question, the lower the summed D score the better adjusted is the *S* considered to be.

It might justifiably be expected that one aspect of the validation problem for an instrument of this type would be to demonstrate that *Ss* with high SC, SA, and IS scores or low D scores are, in fact, better adjusted than their opposites on some independent measure of adjustment. Although there has been a reasonable amount of "validation research" reported for the IAV, the preceding expectation has not always been borne out. Illustratively it is reported (Bills et al., 1951) that scores above the mean on SA relate in one instance to improved adjustment following participation in a student-centered teaching situation (as might be expected from theory) and in another instance to a *higher* incidence of psychotic signs on the Rorschach (quite contrary to theoretical expectations). Subsequently in a larger Rorschach study (Bills, 1953a), a series of 32 hypotheses, dealing with specific test signs which might be expected to differentiate extreme high from extreme low SA scores, was proposed. Parenthetically, 11 of these 32 hypotheses called for no differences between groups, thus having a .95 probability of confirmation. Of the remaining 21 hypotheses, 10 yielded group differences significant at .05 or beyond. Closer examination of these "confirmed predictions," however, indicates that in most instances the high SA scorers (theoretically, the better adjusted group) are significantly *more* maladjusted in terms of the traditional meanings attached to the specific Rorschach signs in question (e.g., less *FC*, and more *CF*, lower $F+\%$, higher dominance of *C* over *M*, etc.). While it may be appropriate to conclude that the high and low SA scorers are differentiable ". . . with respect to certain important personality characteristics as measured by the Rorschach" (Bills, 1953a, p. 38), such an empirical demonstration should not be confused with the validation of underlying theoretical constructs. By and large the findings are opposite in direction to what might be deduced from theory, as well as to other empirical data (LaFon, 1954).

---

[1] Portions of this paper were presented in an address to Div. 8, at the APA meetings in Washington, D. C., August 1958.

[2] Now at the University of Buffalo.

In still another study (Bills, 1953b), the prediction was made that words for which lower IAV ratings were given on a second test administration would also have slower reaction times in a repeat word-association test. Such a prediction is based on the theoretically defensible notion that both lowered ratings and greater latency of association may be symptomatic of conflict or emotionality. The results of the experiment, while statistically significant, were directionally opposite to the prediction. The following statement is advanced in seeking to explain these unanticipated findings:

> . . . the ability to lower a trait rating on re-test may be indicative of a lower degree of defensiveness, and it would be predicted that such a reaction would be accompanied by a decrease in emotionality and, therefore by a decrease in reaction time to a trait in free association (Bills, 1953b, p. 137).

If a lowered SC rating on an IAV retest can be construed as an index of decreased defensiveness, it may be equally defensible to argue that an *initially* low self rating could be due to lowered defensiveness, rather than to conflict or emotionality, as has been assumed. Unless procedures can be specified before the fact, by which we can discriminate the high SC score representing good adjustment from the high SC score representing defensiveness, we are operating within a closed system in which the results of a given experiment, irrespective of their direction, can be interpreted as confirming the underlying theory. Hence, Bills' (1953b) conclusion, on the basis of the experiment cited, that ". . . the index is a valid measure of emotionality and changes in emotionality" (p. 137) appears to be somewhat premature. When an instrument yields significant findings opposite in direction to what might be predicted on the basis of its underlying theory, it may well be discriminatory, but it is not valid in the sense of offering support for the relevant theoretical concepts.

Perhaps the problem would be simplified if, in some consistent way, obtained findings were always contrary to what may be predicted beforehand. But this does not appear to be the case since studies are cited (Bills, 1958) in which high SC scores show up as better adjusted on other measures (e.g., absence of psychosomatic symptoms).

Thus far, most attempts to account for findings which do not support underlying theory in this area (Bills, 1953b; Fey, 1954) have been based upon implications about the vicissitudes of verbal response characteristics in human *S*s. The maladjusted *S*, it is argued, appears well adjusted because of his defensiveness. We may presume that part of the operational meaning of the concept defensiveness, when it is invoked in this context, is the tendency to respond in a socially desirable manner in instances where a true response would place the *S* in an unfavorable light. As yet, the extent to which social desirability stereotypes may account for responses to an inventory such as the IAV, either for specific individuals or for a group of *S*s, is undetermined. It is precisely this type of information which the present research seeks to establish.

Theoretical inconsistencies aside, the brunt of the reported validity data for the IAV (Bills, 1958) is in terms of demonstrated relationships between it and other similarly constructed self-concept measures (e.g., Cowen: 1954, 1956; Omwake, 1954) and/or more traditional paper-and-pencil inventories (Bills, 1958). Such correspondences may, however, indicate little more than constancy in verbal behavior or social desirability stereotypes, and should not per se be taken as a final criterion of validation.

Quite recently, there have been several empirical and theoretical papers directed specifically toward the problem of possible relationships between social desirability and endorsement either of IAV type trait descriptions (Kenny, 1956) or closely related self-descriptive statements from Q-sort pools (Edwards: 1955, 1957; Edwards & Horst, 1953; Kenny, 1956). Most relevant is Kenny's study (1956) in which SD rankings were established for 25 trait descriptive adjectives for a college sample. A second independent sample subsequently rated themselves on these same adjectives both for "real self" and "ideal self." Correlations of .81 and .82 were reported between mean endorsement and SD for these two variables. These data suggest strongly the possibility that a structurally similar formal test instrument such as the IAV might be highly saturated with SD. In essence, the purpose of the present study was to determine the extent to which this was the case. Such an inquiry was made possible by an earlier study in the present series (Cowen, in press) in which mean SD values for a sizable pool of trait descriptive terms, including all 49 adjectives comprising the IAV, were established for a college sample.

Our hypothesis may be summarized as follows:

First, there will be a high positive correlation between both mean SC and mean IS endorsement and SD. Secondarily, we anticipated an inverse relationship between mean SC-IS discrepancy and departure of the mean SD value from the scale midpoint (i.e., the more extreme the SD of an item, the more likely are both SC and IS ratings to be affected by it; hence, the smaller the SC-IS discrepancy) (Kenny, 1956).

Procedure

A sample of 100 college *S*s (59 male and 41 female), independent of the SD normative group (Cowen, in press), completed the standard IAV in group administration. For each IAV adjective mean, SC, IS, and SC-IS discrepancy scores were determined. The mean SC and IS endorsement values were correlated with SD normative values for each trait-descriptive term. Mean D scores for each adjective were correlated with the departure of the SD mean for the midpoint, four, of the SD scale.

For reasons to be noted below it seemed desirable to run a second, cross-validating, sample on a somewhat modified version of the IAV. These latter data were collected and analyzed exactly as the original data had been. The new sample consisted of 23 male and 26 female *S*s.

Results and Discussion

The basic hypotheses. All analyses were run separately for males and females. In each instance, however, the findings were virtually identical for the two sexes; hence, all correlations are reported in terms of the total groups.

Looking only at the data from the initial study we find, in support of our hypotheses, a correlation of .906 ($p > .001$) between SC endorsement and SD. This marked relationship is, in fact, every bit as high as the reliability of the former measure (Bills, 1958). An even higher correlation of .958 ($p > .001$) is observed between IS and SD. On the other hand, the predicted correlation between the SC-IS discrepancy and extremity of SD was not found ($r = - .053$).

Analyzing the actual SD values of the 49 IAV adjectives, it was apparent that they were not normally distributed in terms of the seven-point SD scale used in the normative study. Forty of the 49 terms had SD values ranging from 5.0–6.7, while the remaining nine ranged from 1.1–2.2. This grossly atypical distribution could theoretically have seriously affected the observed correlations. In the instances of the SC and IS correlations, the unusual bipolarity could have operated to inflate the relationships spuriously, while the clustering of 40 of the 49 terms at the positive end of the SD continuum could have had an opposite effect. Additionally, since the deviation of the adjectives from the midpoint of the SD scale was limited to a range of 1.0–2.9, the correlation between this variable and SC-IS discrepancy also could have been restricted artificially.

The second study was undertaken to rule

out the possible attenuating effects of these distributional atypicalities. Forty-eight new adjectives were selected from the original normative study (Cowen, in press) so as to cover the full range of the SD continuum (1.1–6.6). This, of course, also had the effect of increasing the SD deviation scores, the range going from .1–2.9. Although the trait-descriptive terms used for the second study were different from those comprising the original IAV, the instructions, form of administration, and methods of analyses were the same as the ones used with the original sample.

The results of this second study were essentially identical to those obtained with the initial sample. In this instance, the correlations between SC and IS endorsement and SD were .900 ($p > .001$) and .978 ($p > .001$), respectively, while the correlation between SC-IS discrepancy and extremity of SD was −.039. Since these findings occurred notwithstanding the changes in item content as well as their ordering on the SD continuum, it appears that the relationship between SC and IS endorsement and item SD may generalize considerably beyond the IAV test.

Implications. The very high degree of relationship between both SC and IS endorsement and SD raises question as to whether the former two have meaning independent of the latter. Given the known reliability of the self-concept measures, as much of the variance as can be accounted for is accounted for by item SD value. On the basis of these findings, it does not appear to be defensible to refer to the IAV SC and IS scales as indices of self concept and ideal self. These measures are so heavily saturated with SD that it would be more parsimonious to say, when they are found to discriminate empirically, that the observed discriminations are due to differential SD stereotypes rather than to self-conceptual variations. If this position is a sound one, when the interpretive framework is applied to several of the studies given earlier critical review, one might perhaps be tempted to conclude: (*a*) that *S*s markedly influenced by SD stereotypes give more maladjusted Rorschachs (Bills, 1953a) than *S*s who do not respond in terms of such stereotypes, and (*b*) that for those items where *S*s' response became more socially desirable on an IAV retest (Bills, 1953a), word association reaction response latencies also increase. Although direct and independent test of these rewritten generalizations is clearly preferable to the present type of speculation, they may represent a more parsimonious position than do the original generalizations (Bills: 1953a, 1953b).

The SC-IS discrepancy score, on the other hand, appears to be quite independent of the contaminating effects of SD stereotypes. This per se, while it does not establish the validity of this index as a measure of tension or conflict, does, at least indirectly, buttress the findings of other studies purporting to offer such evidence (Cowen, Heilizer, & Axelrod, 1955; Roberts, 1952). However, despite our attempts to maximize the range of SD deviation scores in the second study, mean SC-IS D scores were limited in range from .2–1.4 in both studies. Quite probably this is due to the fact that the IAV measure, as now constituted, is only a five point scale. The narrow range of D scores observed would tend to have a limiting effect upon the magnitude of the correlation coefficient. Perhaps by broadening the range of the IAV rating scales, a somewhat more appropriate test of the second hypothesis could be provided.

Summary

Broadly stated, the purpose of the present study was to assess the degree of relationship between endorsement of items measuring aspects of self regarding attitude and item social desirability.

One hundred *S*s completed the Bills Index of Adjustment and Values. Mean self concept, ideal self, and SC-IS discrepancies were computed for the 49 constituent trait-descriptive terms. The first two of these measures were found, as predicted, to correlate very highly with independently established SD values for these same items. Contrary to prediction, SC-IS discrepancy failed to correlate with deviation of the SD mean from the scale midpoint. Since the IAV adjectives were found to be both skewed and bipolar with respect to their SD values, a new series of 48 adjectives spread out equally across the full range of the SD continuum was selected and given to

a second sample of 49 new *S*s. Using constant methods of test administration and analyses, the results obtained with the second sample were identical to those of the original study.

We have concluded that the self concept and ideal self measures of the IAV are so heavily saturated with SD as to lose meaning independent of the latter variable. Some implications of this basic generalization have been considered.

Received July 18, 1958.

REFERENCES

Berger, E. M. The relationship between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.*, 1952, **47**, 778–782.

Bills, R. E. Rorschach characteristics of persons scoring high and low in acceptance of self. *J. consult. Psychol.*, 1953a, **17**, 36–38.

Bills, R. E. A validation of changes in scores on the index of adjustment and values as measures of changes in emotionality. *J. consult. Psychol.*, 1953b, **17**, 135–138.

Bills, R. E. *Manual for the Index of Adjustment and Values.* Auburn: Alabama Polytech. Inst., 1958.

Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257–261.

Brownfain, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597–606.

Cowen, E. L. The negative self-concept as a personality measure. *J. consult. Psychol.*, 1954, **18**, 138–142.

Cowen, E. L. An investigation of the relationship between two measures of self-regarding attitudes. *J. clin. Psychol.*, 1956, **12**, 156–160.

Cowen, E. L. The social desirability of trait descriptive terms: Preliminary norms and sex differences, in press.

Cowen, E. L., Heilizer, F., & Axelrod, H. S. Self-concept conflict indicators and learning. *J. abnorm. soc. Psychol.*, 1955, **51**, 242–245.

Edwards, A. L. The relationship between judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.*, 1953, **37**, 90–93.

Edwards, A. L. Social desirability and Q sorts. *J. consult. Psychol.*, 1955, **19**, 462.

Edwards, A. L. *The social desirability variable in personality research and assessment.* New York: Dryden, 1957.

Edwards, A. L., & Horst, P. Social desirability as a factor in Q technique. *Educ. psychol. Measmt*, 1953, **13**, 620–625.

Fey, W. F. Acceptance of self and others and its relation to therapy-readiness. *J. clin. Psychol.*, 1954, **10**, 269–271.

LaFon, F. E. Behavior on the Rorschach test and a measure of self acceptance. *Psychol. Monogr.*, 1954, **68**, No. 10 (Whole No. 381).

Kenny, D. T. The influence of social desirability on discrepancy measures between real self and ideal self. *J. consult. Psychol.*, 1956, **20**, 315–318.

Omwake, Katharine T. The relation between acceptance of self and acceptance of others shown by three personality inventories. *J. consult. Psychol.*, 1954, **18**, 443–446.

Phillips, E. L. Attitudes toward self and others: A brief questionnaire report. *J. consult. Psychol.*, 1951, **15**, 79–81.

Roberts, G. E. A study of the validity of the index of adjustment and values. *J. consult. Psychol.*, 1952, **16**, 302–304.

Rogers, C. R. *Client centered therapy.* New York: Houghton Mifflin, 1951.

Snygg, D. S., & Combs, A. W. *Individual behavior: A new frame of reference for psychology.* New York: Harper, 1949.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

TWO QUESTIONS:

A REPLY TO COWEN AND TONGAS

ROBERT E. BILLS

Alabama Polytechnic Institute

The Cowen and Tongas paper seems to have made a significant contribution to understanding a portion of the basis on which self ratings on the Index of Adjustment and Values (IAV) are made. For perspective, though, two questions should be asked.

What Is the Index Used For?

The first question is: What is the index used for?

In the writer's opinion, the emphasis of the Cowen and Tongas paper does not reflect properly what has gone before in the way of research concerns and manner of use of the index. Most of the work on the index has centered around the "acceptance" scores.

The primary reason for designing the IAV (Bills, Vance, & McLean, 1951) was to develop a measure of self-acceptance. To obtain this measure, subjects (*S*s) are asked to answer three questions for themselves for each of 49 trait words. *S*s tell how they see themselves in respect to the trait, how they feel about being this sort of person, and how they would like to be in respect to the trait. The second of these ratings, cumulated for the 49 trait words, has been called acceptance of self. The cumulated difference between the first and third ratings is called the discrepancy score. Discrepancy scores are correlated −.77 (for an *N* of 175) or −.67 (for an *N* of 300) with acceptance of self (Bills et al., 1951; Bills, 1958). The distributions of discrepancy scores and acceptance of self scores are approximately normal (Bills, 1958, pp. 18–19, 22–23, 36–37, 40–41).

For several reasons, including one raised by Cowen and Tongas, we later developed an "others" form of the index (Bills: 1953, 1958) in addition to the "self" form described above. One of the reasons for this was the dilemma posed by high self-acceptance scores being indicators of either high self-acceptance or of defensiveness—a point which Cowen and Tongas also make. To complete the "others" form of the index, an *S* thinks of his peer group and completes the index as he believes the average member of this group would complete it for himself. This procedure seems to have resolved the dilemma (Bills: 1953, 1958). A review of the literature will show the reader that the validation work for the index has centered around the "acceptance" scores from these two forms of the index. In practice, these are the scores commonly used.

This being the case, it does not seem necessary to determine if assumptions one and three as stated in the second paragraph of the Cowen and Tongas paper have been made by this writer. Nor does it seem too important, in the IAV context, to examine the lack of normality of distribution of the social desirability of the index traits. And, we will leave it to the interested reader to determine for himself if "the brunt of the reported validity data for the IAV is in terms of demonstrated relationships between it and other similarily constructed self-concept measures and/or more traditional paper-and-pencil inventories."

Study and use of the IAV have been concerned primarily with the "acceptance" scores from the "self" and "others" forms. Although Cowen and Tongas had data on the self-acceptance scores, they do not report using it. They did use discrepancy scores which are significantly correlated with self-acceptance, and it is important to note that they did not find discrepancy and social desirability significantly correlated.

What Does the Cowen and Tongas Method of Statistical Analysis Yield?

The second question is this: What does the Cowen and Tongas method of statistical analysis yield?

In the writer's opinion, the Cowen and Tongas method is not directly comparable to methods previously used to report IAV research.

The Cowen and Tongas technique for calculating correlations between self concept and social desirability and ideal self and social desirability says, in effect, that the mean of the ratings given by *S*s for each of the 49 IAV trait words is highly correlated with mean social desirability ratings of the traits. This technique is significantly different from that used to calculate the reported reliabilities and the validity of the index. The Cowen and Tongas technique, in effect, ignores the individual variation given by raters. It says that the average ratings for each of the trait words are highly correlated with the average social desirabilities of these traits. This seems to be saying that the average person rates himself as average in social desirability for each of the trait words.

The Cowen and Tongas method assumes (since only the mean of ratings on trait words are used) that the data can be handled as if all *S*s gave the average rating on a trait word. This has the same effect as saying that all *S*s gave the same rating to a trait word as all other *S*s. This is obviously a questionable assumption when information given in the index *Manual* (Bills, 1958) on distribution of ratings for each of the trait words is examined. These data show that considerable variation exists in the ratings for each trait and that the placing of one score (in this case an average rating) on each trait obscures all individual variation, and this is considerable.

But having calculated the correlation between self concept and social desirability and ideal self and social desirability by the above method, the authors switch to another basis for their implication. They compare the above correlations with split-half and test-retest reliabilities obtained from column scores. The reliabilities reported for the IAV were calculated from odd-even scores or column totals in test-retest situations. In other words, Cowen and Tongas arrived at an average rating for each trait word on the index and an average rating of the social desirability of each trait word, paired these, and then proceeded to correlate. They then compared this coefficient with that obtained by totaling columns for each individual and pairing column scores from split halves or from test-retest.

Obviously, these two methods are not directly comparable and the implication cannot be drawn that "Given the known reliability of the self-concept measures, as much of the variance as can be accounted for is accounted for by item SD value." The two techniques yield different results as will be shown. If their's is a valid procedure, it is difficult to see how self concept and ideal self, which are correlated only .55 with each other for an *N* of 300 cases (Bills, 1958, p. 62), could be correlated .91, .96, .90, and .98 with a third variable, social desirability.

Using the formula provided by McNemar (1949, p. 142) and designating the r between self concept and social desirability as r_1, between ideal self and social desirability as r_2, and between self concept and ideal self as r_3 and using .90 for r_1 and .98 for r_2, we can calculate the limiting values for r_3 (that the results would be the same if we used .91 and .96 instead of .90 and .98 can be checked by the reader). In this case, the limits range from .78 to .98. But r_3 for an *N* of 300 cases as has already been stated is .55. Even with as few cases as 49 (the smaller of the groups used by Cowen and Tongas) an r of .55 based on 300 cases is significantly smaller than an r of .78 based on 49 cases at less than the .01 level of confidence.

Received July 18, 1958.

REFERENCES

Bills, R. E. Attributes of successful educational leaders. In R. L. Hopper (Ed.), *Interdisciplinary research in educational administration.* Univer. Kentucky, Bur. Sch. Serv., 1953, **26**, 16–38.

Bills, R. E. *Manual for the Index of Adjustment and Values.* Auburn: Alabama Polytech. Inst., 1958.

Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257–261.

McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

A CONTRIBUTION TO THE "MEANING" OF RORSCHACH'S INKBLOTS VIA THE SEMANTIC DIFFERENTIAL

A. I. RABIN
Michigan State University

The present paper represents an attempt to explore the "stimulus value" of Rorschach's ten inkblots. It is concerned with the global "meaning" which is attributed to them by a group of young normal adults.

Over-all meaning and symbolism of some of the inkblots has been occasionally ascribed intuitively by Rorschach workers. We have undertaken an experimental study of the meaning of these stimuli which is operationally defined by the method called the *semantic differential* (Osgood, Suci, & Tannenbaum, 1957). The rationale for this instrument has been developed and discussed by Osgood (1952) and need not be detailed in the present report.

Procedure

Osgood's semantic differential consists of 50 scales which are defined by polar adjectives, i.e.—adjectives which are opposite in meaning to each other. These poles are separated by seven spaces, one of which may be checked in order to indicate the position of the concept or stimulus rated in relation to the opposite adjective. An example of such a scale is as follows:

Good ___:___:___:___:___:___:___ *bad*

Factor analytic studies of the scales with sizable samples of subjects (*S*s) rating a variety of concepts have yielded three major factors:

I. *Evaluative;* II. *Potency;* and III. *Activity.*

For the purposes of our investigation, 20 scales of the semantic differential were selected by the author. Two criteria governed this selection—relatively high loading on the aforementioned factors and suitability to the material to be judged. The 20 pairs of adjectives employed are listed in Table 1. Factor I (Evaluative) is represented by Items 1, 3, 5, 6, 7, 8, 9, 10, 13, and 16; Factor II (Potency), by Items 2, 4, 12, 15, 18, and 20; and Factor III (Activity), by Items 11, 14, 17, and 19.

The 20-item semantic differential was administered to 66 (28 males and 38 females) undergraduate students in elementary psychology. The 10 Rorschach cards (by means of slides) were projected on a screen one at a time. The *S*s were given three minutes' time to check the items on the semantic differential. Previous tryouts indicated that the time was quite ample even for the slowest *S*s. Standard instructions (Osgood et al., 1957) were given. Thus, a total of 200 check marks (10 stimuli × 20 scales) were obtained for each participant in the experiment.

Statistical Treatment of Data

Although each response (check mark) on the semantic differential has two essential properties—*direction* and *distance*—along the polar continuum (or *quality* and *intensity* of meaning), our main concern for the purposes of this exploratory study was that of *direction* of meaning. Consequently, all the checks, from 1 to 7, were dichotomized, thus using direction only. The number of checks on Space 4, the midpoint of the continuum, was equally divided between the two dichotomized groups of *S*s, i.e., those checking Spaces 1–3 and those checking 5–7. By doing that, the data were prejudiced to some extent in favor of the null hypothesis that there are no differences in the directions of the scale ratings of the Rorschach cards.

The significance of the differences between the number of *S*s checking each direction (the dichotomized data for each scale) was then calculated by means of the chi square technique. Three sets of data were obtained—for the total sample, for the male group, and for the female group separately. Each set of data represents 200 chi square values (10 × 20). A total of 600 chi square figures was thus obtained.[1]

Results

Of the 200 chi square values for the total sample, 112 were found to be significant at the .05 level to beyond the .001 level. The presentation of the levels of significance for all scales and Rorschach cards appears in Table 1.

Using the *direction* of the significance of differences, Rorschach's inkblots may be described by means of the following polar adjectives:

Card I: large, strong, ugly, dirty, distasteful, cruel, unpleasant, ferocious, heavy, cold, thick, active, rough, and rugged

Card II: good, strong, clean, kind, pleasant, happy, peaceful, hazy, and honest

Card III: good, beautiful, clean, tasty, kind, pleasant, happy, peaceful, light, clear, thin, honest, active, smooth, fast, and delicate

Card IV: bad, large, ugly, strong, dirty, distasteful, worthless, cruel, unpleasant, sad, ferocious, heavy, thick, dishonest, active, rough, and rugged

Card V: small, light, clear, thin, active, smooth, fast, and delicate

Card VI: large

Card VII: good, beautiful, weak, clean, tasty, valuable, kind, happy, pleasant, peaceful, light, clear, thin, honest, passive, smooth, and delicate

Card VIII: good, clean, pleasant, happy, peaceful, and hazy

Card IX: good, large, beautiful, strong, clean, tasty, valuable, kind, pleasant, happy, hot, honest, active, and fast

Card X: good, beautiful, clean, pleasant, happy light, thin, active, fast, and delicate

It is interesting to note that not all cards show clear-cut agreement in the characterizations based on the directions in the polar continuum. The meaning of some of the cards, such as IV, VII, III, I, and IX, is delineated in considerable detail, whereas in some of the others, comparatively few significant differentiations are made. This is particularly noted on Card VI where only one significant difference was obtained. A summary of the numbers of significant differences for each card, including a breakdown of the findings with the male and female groups separately, appears in Table 2.

Table 1

Level of Significance (Based on Chi Square) of the Differences in the Incidence of Subjects at Each Polar End in Judging Rorschach's Cards

	I	II	III	IV	V	VI	VII	VIII	IX	X		I	II	III	IV	V	VI	VII	VIII	IX	X
1. good		**	**				***	*	***	*	bad				***						
2. large	***			***		*			***		small					**					
3. beautiful			***				***		***	***	ugly	***			***						
4. strong	***	*		***					***		weak							*			
5. clean		***	***				***	***	***	***	dirty	**			***						
6. tasty			*				*		*		distasteful	***			***						
7. valuable							*		**		worthless				**						
8. kind		***	***				***		*		cruel	**			***						
9. pleasant		***	***				***	*	***	**	unpleasant	*			***						
10. happy		**	***				***	*	***	*	sad				***						
11. ferocious	*			***							peaceful		***	**				***	**		
12. heavy	*			***							light			***		***		***			***
13. clear			***		***		*				hazy		**						*		
14. hot									**		cold	*									
15. thick	*			***							thin			***		*		**			**
16. honest		**	**				***		*		dishonest				**						
17. active	**		***	*	***				***	***	passive							*			
18. rough	***			***							smooth			**		*		***			
19. fast			**		***				***	**	slow										
20. rugged	***			***							delicate			***		*		***			**

* Significant at the .05 level.
** Significant between the .01 and .002 level.
*** Significant beyond the .001 level of confidence.

---

[1] The assistance of James C. Lingoes in processing the data through the Michigan State University digital computer (MISTIC) is hereby gratefully acknowledged.

Table 2

Number of Significant Chi Squares for Each Card for Males, Females, and Combined Sample

	N	I	II	III	IV	V	VI	VII	VIII	IX	X	Total
Males	28	6	4	10	14	7	1	11	2	8	4	67
Females	38	11	7	16	17	4	0	12	2	11	6	86
Combined	66	14	9	16	17	8	1	17	6	14	10	112

It may further be noted, in Table 2, that the number of significant chi square values for the individual sex groups is smaller than for the entire sample. This is readily understood, since the numbers are smaller and a smaller number of chi squares reaches the desired level of significance. Also, the number of significant differences for the male group is smaller than for the female group. This difference is also explained by the smaller number of males when compared with the females. However, what is of particular interest is that no marked sex differences in the direction along the polar continuum of the scales were noted. The males and females showed differences in the same direction. In no instance were there significant differences for males and females in the opposite direction. They never cancelled each other out when combined in the total sample.

Differences in the meanings of the inkblots are further noted in the direction and in the extent to which the three main factors were represented in the evaluation of the cards. By scoring the midpoint of each scale as 0 and the remaining points from +3 to −3, we obtained average factor scores for each card. The plus signs indicate relatively high Evaluation, Potency, and Activity; the minus signs, the opposite (Table 3).

Discussion

In the first place, although each card may have a unique meaning for each individual in the group, there are considerable areas of agreement indicating that the meanings are shared by large numbers of people; i.e., there is a commonality of the meaning reflected in the group results. This finding is consonant with Richards' (1958) observation concerning the individuality of the Rorschach figures. The possible exception may be Card VI on which there is little agreement (one scale out of 20) as to its meaning. To be sure, there is considerable overlapping in the adjectival description of some pairs of cards (such as II and VIII, III and VII); however, in no instance is the resulting profile of any one card exactly like any other in the series. Secondly, the factors into which the scales converge tend to underline further the differences in the meanings; the Evaluative, Potency, and Activity factors have different directions and strengths in the meanings of the several cards. Thirdly, the meanings of the cards seem to be the same for males and females; no significant sex differences with respect to the reactions to the cards were obtained.

These findings must further be related to the available data and hypotheses concerning the Rorschach test. One may address oneself primarily to two major issues which are interrelated—"how do our findings relate to the symbolic meaning attributed to the several cards by Rorschach workers?" and "what information is gleaned concerning the reaction of subjects to the color cards?"—an important theoretical issue in Rorschach theory.

Table 3

Factor Scores for the Ten Rorschach Cards

Factors	I	II	III	IV	V	VI	VII	VIII	IX	X
I (Evaluative)	−.58	+.39	+1.03	− .98	+.25	+.05	+1.09	+.21	+.74	+.38
II (Potency)	+.90	+.03	− .81	+1.69	−.65	−.23	− .88	+.02	+.51	−.41
III (Activity)	+.28	−.17	+ .54	+ .45	+.55	−.17	− .25	−.04	+.76	+.67

Clinicians have especially attributed to three blots of the Rorschach series particular symbolic characteristics. Card IV has been considered as the "father" card, Card VI—"sex," and Card VII, the "mother" card. The symbolism of these cards has also received some support from several experimental studies (Hirschstein & Rabin, 1955; Meer & Singer, 1950; Rosen, 1951). An examination of the meaning of these cards via the semantic differential is, therefore, of particular interest. With reference to Card IV, Bochner and Halpern (1945) stated: ". . . this card embodies something sinister . . . may suggest the father or authority in general." Although, according to the same authors, Card VI ". . . is weighted for sexual implications . . . it is generally . . . the most difficult card to interpret," whereas Card VII is considered to have ". . . a feminine quality, frequently with maternal implications."

If we were to examine the adjectival descriptions of the cards, presented above and based on the semantic differential, we would note some validation of these hypotheses used by Rorschach workers and represented in the descriptions just quoted. The "sinister" and possibly some of the masculine qualities of the card are well reflected in its description as "bad, ugly, strong, dirty, cruel, unpleasant, ferocious, heavy, thick, dishonest, active, rough, and rugged." The "soft" and "feminine" quality of Card VII, on the other hand, becomes expressed via the opposite polar adjectives—"good, beautiful, weak, clean, kind, pleasant, peaceful, thin, passive, smooth, and delicate." It is also interesting to note that on 15 of the semantic differential scales, these two cards differ significantly and appear at the opposite ends of the polar continuum. The obtained profile of the distribution of the adjectives for Card IV is almost a perfect mirror image of that of Card VII. This is the most outstanding and clear-cut difference between any two cards in the series. We have, of course, not validated the "father" and "mother" symbolism, but masculinity vs. femininity, and danger vs. security do seem to be represented by these visual stimuli. Osgood et al. (1957) have briefly reported a study by Smith in which 10 scales were utilized, and the cards were compared with the concepts. Smith's findings did not justify the association between the cards and the parental figures. The selection of the scales may be a determining factor. The detailed findings of this study are not yet available and, therefore, cannot be related to our results.

Whereas Cards IV and VII are "most meaningful" in terms of the semantic differential (17 scales—statistically significant for each), Card VI seems to be "least meaningful" for our *S*s (only one significant difference). It certainly does appear to be "the most difficult to interpret." However, no inkling as to the sexual symbolism was obtained. Maybe the semantic differential scales employed do not tap sex-related meaning. At any rate, within the limitations of the instrument used, it must be stated that this card does not elicit any agreement as to its meaning in our *S*s.

The effects of color in Rorschach's five chromatic cards (II, III, VIII, IX, and X) and the interpretation of the color response have been subjects of considerable controversy as a result of divergent findings in a number of investigations (Ainsworth, 1954). The appearance of color in the Rorschach series has been considered as somehow "disturbing," "displeasing," "startling," and "shocking." The "color shock" phenomenon described originally by Rorschach (1944) was allegedly present primarily in neurotics. Others (Beck, 1946) considered it a much more widespread phenomenon. Our findings with a "normal" student population do not seem to bear out the contention that color is disturbing. Card II, the first one in the series in which color is present, is by and large seen in positive terms. The semantic differential yields such adjectives as "good, strong, clean, pleasant, kind, happy, peaceful, and honest." In view of such descriptive adjectives, it can hardly be called "disturbing" or unpleasant. The remaining color cards (III, VIII, IX, and X) reflect even a greater variety of positively toned qualities and meanings. These results are consonant with Wallen's (1948) experimental findings that "multi-colored cards are better liked in the standard form than when they are rendered achromatically." He adds, however, that "unstable men show more preference for the achromatic version." Perhaps our sample was made up mostly of stable *Ss*. It may also be argued that the "shock," frequently measured by delayed response time, could not be detected in our experimental procedure. But, on the other hand, there seems to be no adequate reason why our *Ss* would rationalize their responses or "cover up" the meaning of, and their reaction to, the cards. To our knowledge there was no special need for the kind of compliance which operates when responses to socially and emotionally conditioned stimuli are required. If interpreted parsimoniously, the results simply indicate that rather positive meanings are attributed to the color cards by a random sample of college students who are presumably largely nonneurotic.

Some of the issues raised above and implied in the findings point to the rich opportunities presented by the semantic differential for the further systematic investigation of the Rorschach technique. Relating the results with the several symbolic cards to meanings obtained with verbal stimuli (e.g., father, mother, etc.), which allegedly they represent, studying differences in the meaning of various cards between normals and clinical groups, observing the relationship between the meaning of cards and the responses produced in the testing situation, and many other approaches can contribute appreciably to the basic understanding of the Rorschach method.

Summary

Sixty-six college students, 28 males and 38 females, checked 20 items of the semantic differential on a seven-point scale for the 10 Rorschach inkblots. The data were dichotomized, and the significance of the differences in the choice of each polar adjective was determined by means of the chi square technique. For the total group, 112 (out of 200) chi square values, significant between the .05 and beyond the .001 level, were obtained. The number of significant chi squares for the males and females treated separately were 67 and 87, respectively. The representation of the three factors—Evaluative, Potency, and Activity—in the obtained meaning of each card was also reported. The results permit the following conclusions:

1. Despite considerable overlap in the meaning of some pairs of cards, no card was described in exactly the same terms as any other.
2. There was considerable range in the number of adjectives, for which significant differences were obtained, which describe any one card. Cards IV and VII are most meaningful, whereas Card VI has practically no commonality of meaning.
3. Cards IV and VII emerged with almost perfectly opposite meanings.
4. The colored cards were considered by the group as pleasing and in, primarily, positive terms.
5. No sex differences in the attribution of meaning to the cards were obtained.

The results were discussed in the light of existing Rorschach theory concerning card symbolism and the effects of color upon Rorschach performance. Suggestions for the further study of the Rorschach by means of the semantic differential were made.

Received July 25, 1958.

REFERENCES

Ainsworth, Mary D. Problems of validation. In B. Klopfer, Mary D. Ainsworth, W. G. Klopper, & R. R. Holt (Eds.), *Developments in the Rorschach Technique*. Vol. I. New York: World Book, 1954. Pp. 405–500.

Beck, S. J. *Rorschach's Test II: A variety of personality pictures*. New York: Grune & Stratton, 1946.

Bochner, Ruth, & Halpern, Florence. *The clinical application of the Rorschach test*. New York: Grune & Stratton, 1945.

Hirschstein, R., & Rabin, A. I. Reactions to Rorschach cards IV and VII as a function of parental availability in childhood. *J. consult. Psychol.*, 1955, **19**, 473–474.

Meer, B., & Singer, J. L. A note on the "father" and "mother" cards in the Rorschach inkblots. *J. consult. Psychol.*, 1950, **14**, 482–484.

Osgood, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 197–237.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.

Richards, T. W. Personal significance of Rorschach figures. *J. proj. Tech.*, 1958, **22**, 97–101.

Rorschach, H. *Psychodiagnostics*. Bern, Switz.: Verlag Hans Huber, 1944.

Rosen, E. Symbolic meanings of the Rorschach cards: A statistical study. *J. clin. Psychol.*, 1951, **7**, 239–244.

Wallen, R. The nature of color shock. *J. abnorm. soc. Psychol.*, 1948, **43**, 346–356.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

WHAT DOES THE RORSCHACH *Z* SCORE REFLECT?

LEON S. OTIS
Johns Hopkins University

When Beck (1933) introduced *Z*, he sought a measure which would reflect the individual's tendency to organize, relate, and abstract aspects of the Rorschach cards into meaningful units. Because the measurement of abstractive and organizing abilities are, presumably, included in many of the current intelligence tests, and because Beck has frequently related *Z* (on an intuitive basis) to high intellectual ability (Beck: 1933, 1949), many authors have assumed that *Z* and intelligence test scores ought to be highly correlated.

The evidence relating *Z* to intelligence test scores or academic achievement, however, has been inconclusive. Wishner (1948) found a significant correlation between *Z* and the Wechsler-Bellevue Fullscale and Subscales, but negative findings between intelligence tests and *Z* have been reported also (Goldfarb, 1945; Jolles, 1947). McCandless (1949) failed to find a significant correlation between academic achievement and *Z*.

The meaning of these contradictory results is difficult to assess, as the subjects of these studies varied greatly as to their mental health (one study used 42 neurotics as its sample), their educational background, their age, and their mental capacity (the sample in one study consisted of 66 feeble-minded children under 18 years of age). Also, these groups differed markedly from the 39 "intellectually very superior" individuals used by Beck (1933) to derive the weights assigned to the different *Z* responses. Beck's sample consisted essentially of 32 individuals in the psychological, medical, and related fields, 5 administrators, a lawyer, and a statistician. Finally, Beck (1949, p. 58) suggested that the "organizational activity" reflected by *Z* is only a *component* of intelligence. Accordingly, the expectation that *Z* ought to correlate highly with intelligence test scores or with academic achievement appears to be based on an overly restrictive definition of intelligence, an unwarranted limitation on the sorts of abilities which *Z*, in fact, does reflect, or both.

If *Z* reflects traits other than those tapped by the usual general intelligence tests, the question is raised: What traits are these and how are they to be identified? If groups which scored high on *Z* and which differed markedly from the general population in other respects could be found, the nature of the correlations between *Z* and the traits which are widely shared by the "high *Z*" group (but not shared by the "low *Z*" group) ought to lead to some hypothesis regarding what *Z* reflects.

The present study is a step in this direction. It compares the *Z* of a group of executives with the *Z* of lower eschelon occupational groups. It seems clear that executives differ markedly in broad intellectual and other personality traits from other employees (see Meyer & Pressel, 1954). This study asks whether the *Z* of executives also differs significantly from nonexecutives.

The sample in this study consisted of 154 of the 157 records of Beck's Speigel group.[1] Three of the records were not available. The records of 36 executives and junior executives (abbreviated, EXEC), 48 skilled workers (SK), 45 semiskilled workers (S-SK), and 25 un-

---

[1] This study was undertaken while the author was serving a clinical internship at the University of Illinois, College of Medicine, Division of Psychiatry. Thanks are due to Alan Rosenwald, director of psychological training at the University of Illinois, for his help and valuable suggestions. The writer is also grateful to S. J. Beck and S. J. Korchin for their cooperation in making the records of Beck's Speigel population available for study.

Table 1

Chi Square Comparisons of the Differences in *Z* Scores Between Executives and Nonexecutives

	At or Below Mdn	Above Mdn	χ^2	*P*
EXEC	13	23	4.86	.03
SK	29	19		
EXEC	13	23	5.46	.02
S–SK	28	17		
EXEC	13	23	7.60	.008
UN–SK	18	7		
SK	23	25	<1.00	N.S.
S–SK	23	22		
SK	24	24	<1.00	N.S.
UN–SK	13	12		
S–SK	22	23	<1.00	N.S.
UN–SK	13	12		

skilled workers (UN-SK), were examined. Membership to these groups was indicated on the protocols.

Table 1 shows that *Z* distinguished the EXEC group from the other groups but did not distinguish between the "nonexecutive" groups. The median scores of the four groups were 27.0 for the EXEC group, 17.0 for the SK group, 16.0 for the S-SK, and 15.0 for the UN-SK, groups.

The wide disparity between the *Z* scores of the EXEC group and the scores of the other three groups is noteworthy. If this finding can be taken at face value, it suggests that *Z* reflects at least some of the characteristics which lead to the top level of the managerial hierarchy.

Of particular interest here, however, is the fact that *Z*, a score which was derived on the basis of examining the Rorschach records of a group of *scholars* and *professional* people, was successful in distinguishing *executives* from other employees. Although differing in many respects (in terms of values, life style, goals, etc.) from each other, scholars and executives obviously share some important traits or characteristics in common. These appear to be reflected in *Z* responses. Perhaps *Z* reflects a rather global trait which might be called "giftedness"; or perhaps *Z* reflects ambition, or intellectual curiosity, or drive. These are some of the traits which scholars and executives might be expected to share. This sort of guesswork is completely unsatisfactory, however. What is needed is a factor analytic study of scholars, executives, and nonscholars and nonexecutives with the Rorschach *Z* score as one of several dimensions of comparison.

Received August 4, 1958.

REFERENCES

Beck, S. J. Configurational tendencies in Rorschach responses. *Amer. J. Psychol.*, 1933, **45**, 433–443.

Beck, S. J. *Rorschach's test. I. Basic processes.* (2nd ed.) New York: Grune & Stratton, 1949.

Goldfarb, W. Organizational activity in the Rorschach examination. *Amer. J. Orthopsychiat.*, 1945, **15**, 525–528.

Jolles, I. The diagnostic implications of Rorschach's test in case studies of mental defectives. *Genet. psychol. Monogr.*, 1947, **37**, 89–198.

McCandless, B. R. The Rorschach as a predictor of academic success. *J. appl. Psychol.*, 1949, **33**, 43–50.

Meyer, H. D., & Pressel, G. L. Personality test scores in the management hierarchy. *J. appl. Psychol.*, 1954, **38**, 73–80.

Wishner, J. Rorschach intellectual indicators in neurotics. *Amer. J. Orthopsychiat.*, 1948, **18**, 265–279.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

BRIEF REPORTS

ECOLOGIC FACTORS IN THE WAIS PICTURE COMPLETION TEST[1]

BERNARD L. BLOOM[2]
Territorial Hospital, Kaneohe, Hawaii

The present study was designed to examine the impact of ecologic factors on the responses to one Wechsler subtest in a situation where pronounced differences were expected and where directional hypotheses could be advanced. Specifically, the Picture Completion subtest (PC) of the Wechsler Adult Intelligence Scale (WAIS) was administered to 67 second- and third-year student nurses in St. Louis, Missouri, and an equal number of student nurses in Hawaii, none of whom had ever been to the mainland. In addition, all subjects were given the WAIS Vocabulary subtest in group form in order to assess the level of linguistic skill.

The following hypotheses were tested: 1. Mainland student nurses will perform higher than Hawaii student nurses on PC Items 11 (stars), 12 (dog tracks), 13 (Florida), and 20 (snow). 2. Hawaii student nurses will score significantly better than student nurses on the mainland on PC Items 8 (peg), 9 (oar lock), 14 (smokestacks), and 15 (crab's leg).

Both groups of student nurses obtained higher average scores and had less variability on the PC and Vocabulary subtests than the standardization sample of equivalent age. The mainland group obtained higher scores than the Hawaii group on both subtests with the difference being at the .01 level on Vocabulary and at between the .05 and .10 levels on PC. The order of presentation on the WAIS test is somewhat more appropriate for the mainland group than for the Hawaii group. The subtest scores correlate .22 with each other in the mainland group and .00 in the Hawaii group.

All of the four items hypothesized to be easier for the Hawaii student nurse group appear to be somewhat easier for this group, and the difference on PC Item 15 (crab's leg) achieves significance at the .02 level. Of the four items hypothesized to be easier for the mainland group, three are easier for this group and the differences on PC Items 12 (dog tracks), and 13 (Florida) are significant at the .01 level. PC Item 11 (stars) is insignificantly easier for the Hawaii group. Thus, of the eight items for which directional differences were predicted, only one comparison failed to lie in this predicted direction.

In addition to suggesting that the PC subtest may be less suitable for use in Hawaii than on the mainland, these results tend to show that specific characteristics of a subject's environment should be taken into account when evaluating test results and that ecologic characteristics may have a predictable and demonstrable relationship to psychological test performance.

Brief Report.
Received January 13, 1959.

---

[1] An extended report of this study may be obtained without charge from Bernard L. Bloom, Territorial Hospital, Kaneohe, Hawaii, or for a fee from the American Documentation Institute. Order Document No. 5894, remitting $1.75 for microfilm or $2.50 for photocopies.

[2] The author wishes to express his appreciation to Ivan N. Mensh and his assistants Judith A. Hiltbrand and Eileen J. Turken, formerly of the Department of Medical Psychology, Washington University School of Medicine, who so kindly provided the student nurse data for the mainland group.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

PERSONALITY IMPLICATIONS OF CIGARETTE SMOKING AMONG COLLEGE STUDENTS[1, 2, 3]

DANIEL S. P. SCHUBERT
University of Chicago

Freud considered smoking an adult extension of autoerotic sucking in childhood, and Fenichel relates oral fixation to the manic-depressive cycle. The hypothesis is suggested that manic-depressive tendencies vary directly with amount of smoking. Vallance's findings (1940) that smokers are less suggestible indicates the second hypothesis that hysteria varies inversely with amount of smoking. Lynn's (1948) conclusion that nonsmokers have better "disciplinary records" at secondary school suggests the additional hypothesis that psychopathic-deviate tendencies vary directly with smoking.

The group form of the Minnesota Multiphasic Personality Inventory was administered to a college group during freshman orientation. The *Ma*, *D*, *Hy*, and *Pd* scores were taken as measures of the tendencies indicated in the hypotheses. The 13 MMPI scale scores (3 validity, 9 clinical, and *Si*) were found to be representative of the northeastern college population from which they were drawn. Of the group, 92 men and 134 women were asked if they smoked cigarettes. Of these, 47 men and 51 women did. This proportion of smokers is also representative of the general northeastern population for this age group (U. S. Department of Commerce, Bureau of Census, 1957).

The results were in the predicted direction on the *Ma*, *Pd*, and *Hy* scales for both sexes and for the females on the *D* scale. The results for males on the *D* scale were not in the predicted direction. Two-tailed t tests showed the significance on the *Ma* and *Pd* scales to be beyond the .001 level for the men and on the *Ma* scale beyond the .01 level for the women. The differences on the *Hy* and *D* scales were not significant. The difference on the *Pd* scale was not significant for the women. Similar t tests on other MMPI scale means were not significant. There was considerable overlap of smokers' scores and nonsmokers' scores on all scales of this study.

Brief Report.
Received July 31, 1958.

REFERENCES

Lynn, R. McG. A study of smokers and non-smokers as related to achievement and various personal characteristics. Unpublished master's thesis, Univer. of North Carolina, 1948.

U. S. Department of Commerce, Bur. of Census. *Statistical abstracts of the United States*, 1957. P. 807.

Vallance, T. R. Suggestibility of smokers and nonsmokers. *Psychol. Rec.*, 1940, **4**, 138–144.

---

[1] An extended report of this study may be obtained without charge from Daniel S. P. Schubert, Department of Psychology, University of Chicago, Chicago 37, Ill., or for a fee from the American Documentation Institute. Order Document No. 5895, remitting $1.25 for microfilm or $1.25 for photocopies.

[2] Paper read at the meeting of the American Psychological Association, Washington, D. C., 1958.

[3] The author is indebted to Desmond Cartwright and Loren Chapman of the University of Chicago for their helpful criticism and suggestions.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

"SOCIAL DESIRABILITY" AND "ANXIETY" VARIABLES IN THE IPAT ANXIETY SCALE[1]

A. W. BENDIG

University of Pittsburgh

Bendig and Hountras (1959) have suggested that Cattell's division of the items in the IPAT Anxiety Scale (Cattell, 1957) into "covert" and "overt" subscales was similar to the "subtle" and "obvious" items in the MMPI and that Cattell's "covert" anxiety items should show a lower correlation with Edwards' Social Desirability Scale (Edwards, 1957) than would the "overt" anxiety items. One difficulty in testing this hypothesis is that 22 of the 39 items on Edwards' SDS are also scored for Taylor's Manifest Anxiety Scale (Edwards, 1957, p. 32) and could be presumed to be more highly loaded with an "anxiety" factor than are the remaining 17 items. In the following research, these two subsets of SDS items were separately correlated with scores from Cattell's scale.

The 40-item Cattell and 39-item Edwards' SDS scales were administered to 238 *S*s (110 men and 128 women) enrolled in educational psychological courses. Seven scores were obtained for each *S*: Covert Anxiety (CA), Overt Anxiety (OA), Difference Score (DS) (CA minus OA), and Total Anxiety (TA) (CA plus OA) from Cattell's scale, and SD-MAS, SD-NonMAS, and SD-Total (SD-MAS plus SD-NonMAS) from Edwards' scale. Correlations among the seven scores were separately computed within each sex group and then averaged, since the men were significantly lower on OA, TA, SD-MAS, and SD-Total.

The correlation between CA and SD-Total (−.56) was significantly lower than the OA and SD-Total correlation (−.70). OA was more highly related to SD-MAS (−.76) than it was to SD-NonMAS (−.48), while CA scores showed similar correlations (−.53 and −.47) with both subsets of SDS items. The TA score correlated higher with SD-MAS (−.71) than it did with SD-NonMAS (−.52). DS correlated significantly with SD-MAS (.36) and with SD-Total (.29), but not with SD-NonMAS (.10). The correlation between CA and OA was .58 and between SD-MAS and SD-NonMAS was .59.

The hypothesized higher correlation of OA and SD-Total was confirmed, but is attributable to a higher correlation of OA with the SD-MAS items. All of the scores, with the exception of DS, seem to measure a general "social desirability" variable which accounts for most of the relationships among the scores. However, OA and SD-MAS appear also to have another variable in common: presumably an "anxiety" component. CA apparently is not a useful measure of "anxiety" except insofar as it operates within the DS score as a suppressor variable for "social desirability."

Brief Report.
Received February 23, 1959.

REFERENCES

Bendig, A. W., & Hountras, P. T. Anxiety, authoritarianism, and student attitude toward departmental control of college instruction. *J. educ. Psychol.*, 1959, in press.

Cattell, R. B. *Handbook for the IPAT Anxiety Scale.* Champaign, Ill.: Institute for Personality and Ability Testing, 1957.

Edwards, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

---

[1] An extended report of this study may be obtained without charge from A. W. Bendig, Dept. of Psychology, University of Pittsburgh, Pittsburgh 13, Pa., or for a fee from the American Documentation Institute. Order Document No. 5919, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 4, 1959

SEX DIFFERENCES IN BODY CONCEPTS[1]

GEORGE CALDEN
Veterans Administration Hospital, Madison, Wisconsin

RICHARD M. LUNDY AND RICHARD J. SCHLAFER
University of Wisconsin

There is little doubt about the importance of bodily and facial attractiveness in American society, yet there has been negligible research in this field. The aim of this study was to compare the degree of satisfaction or dissatisfaction with their body and facial features between males and females. In addition, a comparison was made between the sexes in regard to the attractiveness or unattractiveness of various male body builds.

A "body concept" questionnaire was given to 196 female and 110 male students of an intermediate university psychology course. The form consisted of 13 items pertaining to the *S*'s physical features and to the extent of satisfaction each *S* feels toward these features. Also included was a scale for rating and describing the attractiveness or unattractiveness of seven male body types. The figures, presented on slides, were front view photographs selected from Sheldon's *Atlas of Men*, representing a balanced somatotype and the moderate and extreme forms of endomorphy, mesomorphy, and ectomorphy.

The results indicate that males wish to be three pounds heavier on the average; females seven pounds lighter. All females dissatisfied with their weight wish to weigh less. Only half of the dissatisfied males wished to be lighter. Females, however, are more satisfied with their height than are males. All but two of the dissatisfied males wished to be taller, whereas one half of the dissatisfied females wished to be shorter. Females express less satisfaction with the attractiveness of their bodies. Males wish to have wider shoulders, thicker arms and legs; females prefer smaller hips and waists and thinner arms and legs. Almost half of the females desired larger busts and half of the males desired bigger chests.

Females also are less satisfied with their facial features than are males. *S*s of both sexes single out the nose as the facial feature they most wish to have altered, usually made smaller. Men, however, wished to have more prominent chins and less prominent ears; females preferring bigger eyes, better vision, and a more oval facial shape.

*S*s of both sexes regard the balanced somatotype figure as most attractive and the extreme endomorph as least attractive of the seven male body types. Females, however, viewed the former as more attractive and the latter body type as more repulsive than did the males. Females also were less inclined to view the extreme mesomorph (the "muscle man") as attractive. All of the above differences were significant at the .01 level.

In general, females desire changes from the waist down and wish for smallness and petiteness of body parts (except for bust). Males are dissatisfied with body dimensions from the waist up, desiring bigness of body parts. The greater dislike among females for extreme endomorphic and mesomorphic male physiques may be in part a projection of their personal wish for delicateness in their own bodies.

Brief Report.
Received February 24, 1959.

---

[1] An extended report of this study may be obtained without charge from George Calden, Psychology Service, Veterans Administration Hospital, Madison, Wisconsin, or for a fee from the American Documentation Institute. Order Document No. 5921, remitting $1.75 for microfilm or $2.50 for photocopies.

JOURNAL OF CONSULTING PSYCHOLOGY

October 1959 Vol. 23, No. 5

Contents

ARTICLES

Some Factors in Psychotherapists' Perception of Their Patients: *Norman I. Harway* - 379
Association Tendencies of Groups Differentiated on the Taylor Manifest Anxiety Scale: *E. Philip Trapp and Donald H. Kausler* - 387
An Immediate Memory Modification of the Block Design Task: Relative Performances of Brain-Damaged and Control Subjects: *Alfred B. Heilbrun, Jr.* - 390
Personality and Background Factors in Juvenile Delinquency as Inferred from Questionnaire Responses: *Donald R. Peterson, Herbert C. Quay, and Gordon R. Cameron* 395
The Twisted Pear and the Prediction of Behavior: *Jerome Fisher* - 400
Factor Analyses of Rorschach Scoring Categories and First Response Times in Normals: *Julius Wishner* - 406
Some Manifest Characteristics of Recallers and Nonrecallers of Dreams: *Rosalea A. Schonbar* - 414
Interrelationships Among MMPI Measures of Dissimulation Under Standard and Social Desirability Instructions: *Jerry S. Wiggins* - 419
Intelligence Test Performance and the Delay Function of the Ego: *George Spivack, Murray Levine, and Herbert Sprigle* - 428
Errors in the Diagnosis of Mental Deficiency in Schizophrenia: *Loren J. Chapman and Julian H. Pathman* - 432
Expressed Acceptance of Self and Others by Psychotherapists: *Julian W. Streitfeld* - 435
Discrimination of Female Schizophrenics with Configural Analysis of the MMPI Profile: *William J. Eichman* - 442
Anxiety Indices in Same-Sex Drawings of Psychiatric Patients with High and Low MAS Scores: *Thomas E. Hoyt and Martin R. Baron* - 448
Nomographs for Computing the "Validity" of WISC or Wechsler-Bellevue Short Forms: *Cecil Bridges* - 453
Forced Associations, Symbolism, and Rorschach Constructs: *Joseph F. Rychlak* - 455
A Note on Levin's Report of Findings with Phenobarbital, Promethazine, Chlorpromazine, and Placebo: *Lloyd K. Sines* - 461
A Reply to Sines' Note on a Comparison of the Effects of Phenobarbital, Promethazine, Chlorpromazine, and Placebo Upon Mental Hospital Patients: *Monroe L. Levin* - 463
Absence of Acquiescence Response Set in the Taylor Manifest Anxiety Scale: *Loren J. Chapman and Donald T. Campbell* - 465

BRIEF REPORTS

Validation of Doppelt's WAIS Short Form with a Clinical Population: *Hugh Clayton and Dan Payne* - 467
The Validity of a Self-Report Measure of Anxiety as a Function of the Time Interval Covered by the Instructions: *Barclay Martin* - 468
Manifest Anxiety and Performance on Problem Solving Tasks: *Durganand Sinha and Tripat Raj Singh* - 469
PSYCHOLOGICAL TEST REVIEWS - 470

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.

2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.

3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.

4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual*. Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

> [1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual*.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

SOME FACTORS IN PSYCHOTHERAPISTS' PERCEPTION OF THEIR PATIENTS[1]

NORMAN I. HARWAY
School of Medicine and Dentistry, University of Rochester

The psychotherapist, during the course of his contacts with a patient, attains a degree of knowledge of his patient. The understanding that the therapist develops with regard to his patient is not, however, merely a matter of extent or quantity of information. Both in the internal experience of the therapist and in his expressions of his knowledge it is evident that different degrees of certainty adhere to the varying things that the therapist knows about his patient. Some statements about the patient are offered with little hesitation or doubt, whereas others are formulated tentatively and expressed with many qualifications.

There has been much discussion of the adequacy of clinical judgment. These discussions for the most part have been concerned with whether statistical or clinical methods lead to a greater number of accurate predictions of patient behavior (Meehl, 1954). Ability to predict is but one criterion of knowledge, however. Awareness of what one does or does not know regarding a person would also be a reflection of knowledge as would internal consistency in the statements expressing one's understanding of a person.

A number of studies in recent years have been concerned with the ability of one individual to predict the behavior of another individual. In connection with psychotherapy, Kelly and Fiske (1951) used the therapist's ability to accurately predict patient response as one measure of therapeutic competence. The therapist's estimation of the similarity between himself and his patient, or of the similarity between the therapist's "ideal" and the patient's self-descriptions, has also been suggested as relating to competence (Fiedler, 1951; Kelly & Fiske, 1951). Dymond (1953), who used the predictive method as a measure of empathy, has shown that counselors' predictions of client responses are not primarily a function of the counselors' stereotype of the average client.

The therapist's statements regarding the patient derive from many sources. Among those which may be posited are the patient himself as a stimulus, some general attribute or capacity of the therapist as a perceiver or knower of other people, the personal qualities of the therapist as reflected in his conflicts, desires, attitudes, and needs, and the theoretical schema which the therapist utilizes in observing and thinking about the patient. These form a complex structure which relates to the therapist's understanding, empathy, and reactions to the patient.

The present study is an attempt to explore some aspects of this structure. The questions investigated are: (*a*) Is the therapist aware of what he knows and does not know about the patient? In terms of the operations here involved we may phrase the question: Is the therapist's confidence in his predictions of a patient's responses to a personality questionnaire related to the accuracy of these predictions? (*b*) Is there any relation between the therapist's knowledge of specific behaviors of his patient and his description of the patient in more abstract terms? This may be examined by a comparison of a personality profile derived from individual item predictions with a profile constructed in terms of the therapist's more global perceptions of the patient. (*c*) Does the therapist's orientation relate to the prediction he makes for the patient? In part, the therapist's orientation or viewpoint is reflected in the particular goals he sets for a patient and in the relative im-

---

[1] This investigation was supported in part by a grant from the Foundations Fund for Research in Psychiatry.

portance he assigns these goals. It is this somewhat specific feature of the therapist's frame of reference that is tapped in the present instance. (*d*) To what extent do the therapist's own characteristics influence his perception of the patient and the goals that he sets for the patient?

While the problem falls under the general rubric of interpersonal perception, the psychotherapeutic setting was chosen because of the intensity of the relationship that develops and the role assignments in which it is the duty of one person—the therapist—to know, to understand, to empathize with the other. The therapeutic procedure is designed in many ways to maximize knowledge of the patient.

Procedure

Nine psychotherapists participated in the present study. The attempt was to obtain a reasonably homogeneous therapist group. In the current instance the nine were either senior psychiatric residents or junior staff members in psychiatry, all of whom had from two to four years' training in psychiatry at the same institution. While their background was in many ways eclectic, the major frame of reference might best be categorized as psychoanalytically oriented. All, at the time of the study, were treating patients in the outpatient department of a university hospital, none were fully trained psychoanalysts, nor were they using the formal psychoanalytic method, and all were working under the supervision of a tutor with whom they met periodically and frequently to discuss their psychotherapeutic activities.[2]

The instrument used was the Edwards Personal Preference Schedule (EPPS) (1954). This is a 225-item personality questionnaire based on the theory of Murray (1938). The test is scored in terms of 15 manifest needs: achievement, deference, order, exhibition, autonomy, affiliation, intraception, succorance, dominance, abasement, nurturance, change, endurance, heterosexuality, aggression. The test is designed to minimize the influence of social desirability as a response determiner. Statements representative of each need area are paired twice with each of the other manifest needs. For each item (pair of statements) the subject (*S*) records which of the two statements is more characteristic of what he likes or how he feels. A measure of *test consistency* and a measure of *profile stability* can be obtained as well as the primary need scores.

Each of the nine therapists filled out the EPPS with the usual instructions, that is, as he felt the various items were applicable to himself. Shortly after this, the writer met individually with each therapist and reviewed with him the patients under his care. While an attempt was made to select a patient from the caseload of each therapist who would be homogeneous with respect to such variables as age, sex, diagnosis, length of treatment, and type of treatment with the patients selected for the other therapists, only limited success was achieved. The essential characteristics of the patients are presented in Table 1. The age range is 17 to 35; six are male, three female. All patients had been seen by their particular therapist for at least 20 hours.

At least six weeks were permitted to elapse before the second step in the study. Arrangements were made with each therapist and patient so that at the end of a specified therapy hour the patient immediately went to the psychology offices where the EPPS was administered. At the same time, back in his office, the therapist attempted to answer the EPPS as he thought the patient was doing it.[3] Additional tasks were set for the therapist. As he responded to each item as he thought the

---

[2] The writer wishes to express his appreciation to the psychiatrists who gave so generously of their time and to the patients who participated in taking the personality inventory. I am indebted to the participants in the Research Conference of the Department of Psychiatry for their counsel.

[3] Other writers have made the distinction between *prediction*, where one states in advance what another person will do, and *postdiction*, where one says what the other individual has already done though knowledge of the other's performance has been withheld. In this study, then, we are dealing with *codiction*. The term *codiction* will be used when referring to our procedure to emphasize that the therapist was aware of both the setting in which the patient was responding and the immediate mood of the patient. When referring to the work of others or to the class of procedures relating to one person saying what another has done, is doing, or will do, I will use the more usual term—*prediction*.

Table 1

Description of Patient

Patient	Sex	Age	Education[a]	Interviews with Therapist	Diagnosis
1	M	21	15	21	Character trait disturbance
2	M	19	9	85	Adjustment reaction of adolescence
3	M	25	18	65	Character trait disturbance
4	M	17	11	37	Obsessive-compulsive neurosis
5	M	22	12	75	Schizophrenic reaction, simple
6	M	23	16	48	Adjustment reaction of adolescence
7	F	33	12	100	Hysterical character disorder
8	F	35	14	30	Conversion reaction
9	F	20	15	121	Adjustment reaction of adolescence

[a] Number of years of formal schooling.

patient was responding, he was asked to indicate on a five-point scale the degree of confidence that he had that his statement for the particular item would match the patient's response to that item. When he completed this task the therapist took a sheet which had listed on it the 15 need areas for which the EPPS is scored. Each area was defined as in the EPPS manual. The therapist rank ordered the 15 areas in terms of his opinion as to their dominance for the patient at that time. Assuming the validity of the EPPS the therapist was, in essence, making a crude psychogram of the patient's personality.

After a lapse of at least six weeks the therapist was again given the EPPS; this time with the instruction to complete the form as he would like to have his patient complete it at the end of the therapeutic process. At the same time, the therapist was to rate on a five-point scale how important it was to him that at the end of treatment his patient complete each item in this particular way.

To recapitulate, for each of the 15 need areas we have seven scores based on: (*a*) patient's performance on questionnaire, (*b*) therapist's performance on questionnaire, (*c*) therapist's codiction of patient's performance, (*d*) therapist's confidence rating for the accuracy of each of his codictions, (*e*) rank order psychogram by therapist, (*f*) therapist's statement of what he would eventually want patient to be like, (*g*) importance ratings by therapist for each item as he does the immediately preceding task.

Results

We have treated each patient-therapist pair as if they constituted a separate experiment. Thus, the nine pairs can be taken as nine replications of the same experiment. This enables each pair to act as its own control. Our concern, then, is to compare for each therapist those areas for which he shows maximum knowledge with those for which he shows less knowledge. We have then combined the results of these nine replications on the assumption that these might be considered as successive samples from a population of similarly qualified therapists. The nine therapists represent nine-tenths of the population of therapists meeting the description given above who were available to us locally.

The normative sample on which the EPPS was standardized consisted of college men and women between the ages of 15 to 59, with the majority being at the lower end of the age range. To test the adequacy of the test for our sample, the measures of test consistency and profile stability were computed. Test consistency is measured by repeating 15 items in the inventory; the score is the number of items answered in identical fashion on both occasions. Profile stability is evaluated by correlating the sum of half the items contributing to a need score with the sum of the alternate half. The correlation is across the 15 personality variables for a single *S*. The data are presented in Table 2. Edwards (1954) suggests that if an *S* obtains a consistency score of less than nine the validity of that particular test may be questioned (p. 7). Neither the patients nor the therapists, regardless of the instruction set used, show a lack of consistency by this standard. Further, the profile stability coefficients compare favorably with those reported in the test manual. The median correlation reported there is approximately .71. It seems proper then to use this test with a patient population. It is also reasonable to conclude that the various instructional sets did not invalidate the use of the questionnaire.

Assuming the equal likelihood of a positive or a negative relationship between the variables in which we are interested, we find that by the binomial method we would expect all nine replications to come out in the same direction twice in a thousand times and eight out of nine to fall in the same direction two times in one hundred. However, as we set forth no directional hypotheses, a two-tailed test seemed appropriate and the respective values would then be .004 and .04. Eight out of nine in the same direction with a p value of .04 is taken as constituting a statistically significant finding.

The therapist's score for each need area when he attempted to codict the patient's responses was compared by means of Q correlation with the scores derived from the patient's performance on the EPPS. The nine correlations, one for each patient-therapist pair, are presented in Table 3, Column 1. It will be noted that eight of the nine correlations are positive, one falling at the .03 probability level. If we compare the rank order psychogram as drawn up by the therapist, with the relative ordering of the areas as derived from the patient's responses (Column 2), we find that all nine correlations suggest a positive relationship. The negative sign in this instance results from a rank of one representing the most dominant area and a rank of 15 representing the least dominant area in the therapist's ranking. These were correlated with the actual scores for the 15 need areas based on the patient's responses where a high score represents the relative dominance of an area. Here, again, one of the individual correlations is significant at the .02 level. It is possible, however, that one of these sets of correlations is a statistical artifact. There is a high degree of relationship between the therapists' codictions at the two levels—the sum-of-items score and the more global rank ordering. Table 3, Column 3, shows that all nine coefficients imply a positive relationship, and all are significant at the .02 level or beyond. If one of these two levels correlates with the patient's psychogram, then the fact that the other does so also is a matter of statistics rather than psychology. It is not possible to extract from these data which two of the three sets of correlations (Columns 1, 2, 3) are independent. Are the therapists equally good codictors at both levels and therefore the two levels turn out to be correlated? Or does the relationship between codiction and

Table 2

Test Consistency and Profile Stability Measures

	Patient		Therapist		Codiction		Goals	
Patient	Consistency	Stability	Consistency	Stability	Consistency	Stability	Consistency	Stability
1	12	.96	12	.82	13	.87	13	.87
2	11	.73	14	.92	11	.93	12	.79
3	15	.84	10	.89	13	.88	11	.80
4	13	.74	14	.91	14	.93	15	.97
5	14	.88	13	.91	12	.84	14	.92
6	9	.92	14	.78	12	.82	12	.84
7	10	.95	13	.83	14	.84	11	.84
8	10	.92	11	.62	12	.77	13	.70
9	13	.93	11	.84	14	.94	11	.89
Median		.92		.84		.87		.84

Table 3

Intercorrelations

$(N = 15)$

	1	2[a]	3[a]	4	5	6	7
Patient	Patient-Codiction	Patient-Rank Order	Codiction-Rank Order	Patient-Therapist	Codiction-Therapist	Accuracy-Confidence	Accuracy-Consistency
1	.38	−.35	−.78*	.49	.44	.20	.14
2	.37	−.04	−.85*	.38	−.13	.14	.33
3	.23	−.03	−.71*	−.42	.14	.19	−.32
4	.36	−.06	−.60*	.12	.15	.38	.37
5	.28	−.27	−.83*	.24	−.07	.70*	−.25
6	.18	−.40	−.75*	.63*	.30	.30	−.15
7	.38	−.31	−.91*	.00	.20	.31	−.25
8	.55*	−.57*	−.72*	.00	.47	−.28	.00
9	.00	−.16	−.76*	−.01	.39	.19	−.19

[a] Negative correlation signifies positive relationship.
* $p \leqq .05$.

patient response exist at only one level, but since the therapists are consistent at different levels there is also a positive relationship at the second level?

The size of the correlations, however, suggests that the individual therapists are not highly efficient predictors. Our results are comparable to Dymond's (1953); she reports correlations ranging from .05 to .84, with a median of .41 for counselor predictions of client responses at the end of treatment. Column 1, Table 3, shows a range from .00 to .55, with a median of .36. Reanalyzing our data by the more usual procedure of an item-by-item rather than need-area score basis and assuming that a random response would lead to half the items being accurately codicted, we find that in all nine instances the therapist had more accurate than inaccurate codictions, seven of the nine chi square values are significant beyond the .05 level, and the combined chi square with nine degrees of freedom has a $p < .001$.

To analyze another triad which may lead to a spurious evaluation of accuracy of codiction: there is no consistent relationship between the relative strength of patient's needs and those of the therapist (Table 3, Column 4) nor is there a consistent relationship between the relative strength of the needs for the therapist himself and the relative strength that he ascribes to them for the patient (Column 5).

To test whether the therapist knew what he knew, the confidence he expressed in his codictions for a need area was compared with his accuracy in codicting for that area of the patient's personality. Correlating the sum of the confidence ratings for each area with the number of accurately codicted items in that area we find that eight of the nine correlations are positive, one being significant beyond the .01 level (Table 3, Column 6). It seems likely, then, that the therapist's sense of certainty or doubt is related to his actual knowledge. Here, again, the individual correlations are not very high. However, comparing mean confidence rating of accurately codicted items with the mean confidence rating of the inaccurate items, the mean rating of the former is greater for all nine therapists, four of the nine t tests are significant at the .05 level or better, and the combined probability of the nine t tests is less than .001. The fact of greater or lesser consistency between the two levels of prediction by the therapist for a given area did not seem to relate to the accuracy in predicting for that area of the patient's personality (Column 7).

The final task set for the therapist was to indicate how he would like the patient to answer the statements on the EPPS at the completion of treatment and how important he felt it to be that the patient answer the particular response in the particular way. If we compare the scores for each need as derived from this response set with the importance assigned to the items in the various need areas (Table 4, Column 1), eight of the nine correlations are positive, three being significant beyond the .03 level. It would seem, then, that those areas in which the therapist would like his patient to score high are the areas which are most important for the therapist. Neither the goal scores nor the importance ratings show a consistent pattern over the nine therapists with the codictions (Columns 2 and 4) nor with the rank ordering of the areas by the therapist (Columns 3 and 5). In other words, they do not seem to have influenced the therapist's predictions for the patient, nor his perception of the patient at the time of the study. The goal scores, however, are related to the therapist's own responses to the questionnaire (Column 6); two correlations are significant beyond the .03 level. The indication here is that the therapist would like the patient to become more like the therapist as a result of treatment. The importance of an area for the therapist did not relate to his accuracy of codicting that area (Column 7).

Discussion

It has been suggested that accurate prediction may be a function of factors other than knowledge of the predictee (Cronbach, 1955; Gage & Cronbach, 1955). In certain instances it may be the result of some general response set. The EPPS, in which the social desirability of the various subscales is controlled, was chosen to eliminate the likelihood of a general desirability or social acceptability set determining codictions. In certain instances predictive accuracy may be enhanced by "projection." In terms of the operations involved, if the predictor projects—that is, answers for the predictee as for himself—then he may be accurate not through any knowledge of the patient but because the predictee "accidentally" happens to be like himself. In the present data there is no significant relationship between the need patterns of patient and therapist nor between the therapist's own pattern and the one he codicts for the patient.

In requiring a therapist to make a set number of predictions about his patient to obtain an over-all measure of the therapist's understanding, the assumption is made that equally relevant information is available to all predictors and that equally relevant information is available regarding the various predictions to be made. The knowingness of the therapists, however, resides as much in their awareness of their relative understanding of different aspects of their patients as well as in their ability to codict what their patient will do on a personality questionnaire. The positive correlation between the item codiction profile and the rank order profile suggests an internal consistency to the therapist's conceptualization of the patient. The consistency between the formulations at the two levels is

Table 4

Intercorrelations

$(N = 15)$

	1	2	3[a]	4	5[a]	6	7
Patient	Goals-Importance	Goals-Codiction	Goals-Rank Order	Importance-Codiction	Importance-Rank Order	Goals-Therapist	Importance-Accuracy
1	.31	−.11	.42	−.32	.17	.42	−.37
2	.05	.11	.05	−.26	.12	.56*	.00
3	.47	−.32	.45	.29	−.23	.30	.41
4	−.06	−.08	.50	.40	−.36	.73*	.51
5	.56*	.57*	−.51	.52*	−.46	.17	−.22
6	.29	.39	−.23	.22	−.24	.47	−.22
7	.81*	−.10	.08	−.17	.23	.20	.13
8	.12	.31	.11	.36	−.27	.31	.16
9	.87*	−.27	.54*	−.38	.68*	.25	−.25

[a] Negative correlation signifies positive relationship.
* $p \leqq .05$.

not related, however, to the accuracy of codiction at either level.

The therapist's knowledge or understanding of his patient has been treated here as a variable rather than as a constant process. Those studies utilizing predictive accuracy or assumed similarity (the similarity between the therapist's predictions for the patient and his own responses to the task) as measures of therapeutic competence, empathic ability, or a general tendency to have a warmer feeling and liking for one's patient focus on these measures as indices of constant processes within the therapist, that is, as being primarily a general attribute or capacity of the therapist (Cronbach, 1955). Our assumption has been that understanding is not only a function of the patient (that is the stimulus object) as well as of the therapist and the setting in which he knows the patient, but that understanding is variable with respect to different aspects of the stimulus object. In the present instance, then, we have compared different indicators of the therapist's knowledge with respect to one stimulus object and have investigated the relationship of understanding to the viewpoint or frame of reference of the therapist.

Though we have used the terms viewpoint, frame of reference, and orientation somewhat interchangeably, the thing we are interested in is the theory which the therapist utilizes. Our approach here has been indirect and admittedly too limited. The problem of investigating the higher order abstractions of the clinician in a manner which will enable direct comparison with more concrete statements has been a stumbling block (Cartwright & French, 1939). The consistency between the therapists' formulations at the two levels obtained in the present study provides some data on this matter. Unfortunately, Murray's system on which the EPPS is based, though closely related to psychoanalytic thinking, is not identical with the frame of reference of our therapists, and many of our therapists probably had not yet achieved a clear, consistent, conscious theoretical position. Our findings are also limited by the level of training of the therapists.

The discrepancies that exist between analyses based on summing the individual items into need scores and those based on the items themselves (cf. agreement between patient response and codiction and between confidence rating and accuracy) also may be considered in this context. The therapists showed a high degree of accuracy in codicting the individual items, yet when the codictions are summed by Edwards' scoring system the relationship to the profile based on the patients' responses is not very impressive. Perhaps the items do not sum in the same way for our group of therapists as they do by the scoring procedures. The high correlation between the therapists' rank ordering of the needs as they apply to the patients and the profiles derived from the individual codictions suggests that the therapists when forced to use the 15 manifest needs are in agreement with the scoring system as to which items relate to a given need. But these 15 needs are not the dimensions underlying the therapists' codictions. We may speculate that the accuracy on the individual items results from a different frame of reference than that on which the questionnaire is based. Perhaps the more fundamental question is whether the scored dimensions of the test are the same as the personality dimensions of the patients.

The lack of relation between the goal responses or the importance ratings and the codictions made by the therapist indicates that the therapist's orientation is not related to his knowledge of the patient. While it might be expected that the goal-set responses would differ from the codictions, that is, that the therapist would want the patient to become something different than he now perceives him as being, there is no clear negative relationship. One might also have expected that those areas which were considered more important by the therapist would be those of which he was most aware and thus those for which he could codict most accurately. This was not the case. Rather, the goals appear to be closely related to what the therapist himself is like. Correlation between the goal-set responses and importance scores leads to the possible conclusion that therapists are more concerned with those attributes they want the patient to achieve than with those attributes they would like the patient to minimize. To combine these last two statements, we may push our conclusions a bit further and say that it is more important to the therapist that the patient exhibit the same dominant needs as he (the therapist) does than that the patient minimize certain aspects of his own current functioning. The focus is on what one should be rather than on what one should not be.

Summary

The structure of psychotherapists' understanding of their patients was explored in a study of nine therapist-patient pairs. The assumption was that therapists have different degrees of understanding of the various needs of their patients. From the patients' responses to the Edwards Personal Preference Schedule and the therapists' utilization of the schedule under a variety of instructional sets, the following conclusions are suggested: (*a*) therapists are aware of what they know and do not know about their patients, (*b*) therapists are consistent in their evaluation of their patients at a more concrete and at a more abstract level of formulation, (*c*) the goals that a therapist sets for the patient are related to the therapist's personality structure, (*d*) in setting goals for his patient the therapist is more concerned with what the patient should be like than with what he should not be like. The conclusions are offered tentatively because of the small size and limited nature of the sample both with respect to the therapist and patient populations and the areas of personality investigated.

Received September 17, 1958.

REFERENCES

Cartwright, D., & French, J. R. P., Jr. The reliability of life history studies. *Character & Pers.*, 1939, **8**, 110–119.

Cronbach, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

Dymond, Rosalind. Can clinicians predict individual behavior? *J. Pers.*, 1953, **22**, 151–161.

Edwards, A. L. *Personal Preference Schedule.* New York: Psychological Corp., 1954.

Fiedler, F. E. A method of objective quantification of certain countertransference attitudes. *J. clin. Psychol.*, 1951, **7**, 101–107.

Gage, N. L., & Cronbach, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.

Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. Michigan Press, 1951.

Meehl, P. *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: Univer. Minnesota Press, 1954.

Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

ASSOCIATION TENDENCIES OF GROUPS DIFFERENTIATED ON THE TAYLOR MANIFEST ANXIETY SCALE[1]

E. PHILIP TRAPP AND DONALD H. KAUSLER
University of Arkansas

Theoretical studies on the Taylor Manifest Anxiety Scale (MAS) have mainly been concerned with the relationship of MAS scores to drive level. The likely possibility that MAS scores may also reflect differences on other dimensions, such as differential habit or associative tendencies, has been frequently propounded (e.g., Child, 1954; Farber, 1955; Lazarus, Deese, & Osler, 1952), although little empirical evidence has actually been adduced. Farber (1955), in a recent review, aptly summarizes the untenable status of this particular property of the MAS in stating: ". . . it is not as yet entirely certain what these habitual differences may be, apart from the trivial observation that they consist, at least in part, in the kinds of verbal responses given on the test itself" (p. 324).

The present study was designed to investigate the nature of some of the associative tendencies presumably related to MAS scores. Since the MAS was developed directly from items of the MMPI purporting to reflect manifest symptoms of the anxiety state, it should follow that high scorers on the MAS would produce associations with attributes similar to those expected in a population of clinically anxious subjects (*S*s). One commonly ascribed attribute of the associations of clinically anxious *S*s is the relatively high proportion of negative tones or implications. Therefore, the approach taken in this study was to compare the proportionate frequencies of negatively toned associations elicited from high and low scorers on the MAS to some neutral stimulus task. The selected stimulus task was a list of nonsense syllables. Predicting that MAS scores would reflect differential associative patterns, the hypothesis was stated as follows: *S*s scoring at the upper extreme on the MAS will reflect significantly more negatively toned associations to a list of nonsense syllables than *S*s scoring at the lower extreme on the MAS.

Method

Subjects. The *S*s consisted of 43 students, male and female, in general psychology classes at the University of Arkansas. They represented the two extremes of scorers on the MAS from a group of over 300 students. The *S*s were tested during regular class periods.

Procedure. The 320 nonsense syllables in Hull's (1933) list were prepared for group administration in a booklet containing 20 pages, with 16 syllables to the page, arranged in random order. The *S*s were informed that they were participating in a study of the association value of nonsense syllables. They were instructed to look at each syllable briefly and, if the syllable aroused any particular thought or idea, they were to write it down on the separate answer sheet. The use of a template restricted their attention to one syllable at a time.

The MAS was given some two weeks to a month later, disguised as a biographical inventory, by a different *E*. A short form of the inventory, consisting of the 50 anxiety keyed items and 40 buffer items, was used. The *S*s were informed that *E* was collecting bio-

---

[1] This study was supported by a grant from the College of Arts and Science Research Fund, University of Arkansas. Edward Mequet assisted in collecting the data.

graphical information about students at Arkansas to compare with students elsewhere.

Results and Discussion

To simplify the statistical treatment of the data and control for possible group differences in total number of associations, 25 associations were randomly selected from the answer sheets of the *S*s scoring at the two extremes on the MAS. The high scorers (Group HA, N of 21 scored 28 or above on the scale; the low scorers (Group LA, N of 22) scored 6 or below on the scale. The groups consisted of approximately the top 10% and bottom 10%, respectively, of the total distribution of scores. Each set of 25 associations was then typed on a separate answer sheet and coded with respect to group identity.

Two judges, staff psychologists at the University of Arkansas, independently evaluated each association, classifying it as positively toned, negatively toned, or neutrally toned. A total of 1,055 judgments, consequently, were made. Percentage of agreement between the two judges was 92.

A summary of the means and standard deviations of the positively toned and negatively toned associations for the two groups is presented in Table 1.

The hypothesis that high scorers on the MAS would reflect significantly more negative associations than low scorers was confirmed by the t test beyond the .001 level ($t = 5.75$; $p < .001$). Incidentally, no significant difference was found between the groups with respect to number of positively toned associations ($t = 1.09$; $p > .05$). The preponderance of associations made by both groups was of the neutrally toned variety. This would be expected from the nature of the task.

Table 1

Summary Data of Number of Associations to Nonsense Syllables

Group	Negatively Toned Associations		Positively Toned Associations	
	Mean	*SD*	Mean	*SD*
HA (High Anxious) ($N = 21$)	6.29	2.19	2.10	1.26
LA (Low Anxious) ($N = 22$)	2.91	1.74	2.55	1.47

A check was made on the sex distribution of the two groups, and no appreciable difference was obtained. Group HA consisted of 13 males and 8 females, and Group LA consisted of 16 males and 6 females. Hence, the difference in the obtained associations is unlikely an artifact attributable to the sex variable.

The findings thus support the theoretical position that the MAS does differentiate *S*s on associative factors as well as the more frequently investigated nonassociative factors. In other words, evidence is offered that MAS scores reflect, in addition to the energizing component of a motivational state, a directing component, which some writers (e.g., Dollard & Miller, 1950; Farber, 1955; Hull, 1943; Shaffer, 1936) feel is an essential requirement of a motivational variable.

Summary

The present study was an investigation of some of the associative tendencies that may be reflected in MAS scores. The response patterns of 21 high scorers on the MAS and 22 low scorers were compared in terms of the frequencies of negatively aroused associations to a list of nonsense syllables. The hypothesis that high scores on the MAS would produce a significantly greater proportion of negatively toned associations was confirmed beyond the .001 level ($t = 5.75$; $p < .001$). The groups reflected, on the other hand, approximately the same proportion of positively toned associations. The sex ratio was nearly the same in both groups, thereby eliminating the possibility of the sex variable confounding the results. Hence, the most probable explanation is that MAS scores do reflect differential associative tendencies.

Received August 1, 1958.

REFERENCES

Child, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–171.

Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

Farber, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.

Hull, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.

Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

Lazarus, R. J., Deese, J., & Osler, Sonia F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

Shaffer, L. F. *The psychology of adjustment*. New York: Houghton Mifflin, 1936.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

AN IMMEDIATE MEMORY MODIFICATION OF THE BLOCK DESIGN TASK:

RELATIVE PERFORMANCES OF BRAIN-DAMAGED AND CONTROL SUBJECTS[1]

ALFRED B. HEILBRUN, JR.
State University of Iowa

The standard block design procedure requires a subject (*S*) to construct block patterns with the design card as a continual reference source. Under these conditions the task has been demonstrated to be differentially difficult for *S*s with brain damage when compared with various types of control groups without brain damage (Aita, Armitage, Reitan, & Rabinowitz, 1947; Allen: 1947, 1948; Boyd, 1949; Goldman, Greenblatt, & Coon, 1946; Greenblatt, Goldman, & Coon, 1946; Hécaen de Ajuriaguerra & Massonet, 1951; Heilbrun, 1954; Lidz, Gay, & Tietze, 1942; McFie & Piercy, 1952; McFie, Piercy, & Zangwill, 1950; Milner, 1952; Paterson & Zangwill, 1944).

There have been attempts to modify the standard block design procedure to increase its sensitivity to pathological brain states. Grassi (1947, 1953) in his Block Substitution Test requires that an *S* copy a set of blocks rather than a drawing. In addition, the test includes using different colors from those of the model, and *S* is required to copy all six sides of the model. The Block Design Rotation Test devised by Shapiro (1951, 1952, 1953) is scored by the amount of rotation of the completed block design. Both investigators report successful discrimination between brain-damaged and control groups, but in neither case was there an adequate comparison between the revised and standard block design procedures. This would seem necessary before procedural revision could be advanced as a diagnostic improvement.

The present study evaluated a modification of the standard block design procedure by requiring *S* to construct the designs from memory immediately following exposure of the stimulus card. The modified procedure was based on the hypothesis that a combination of two psychological processes, each independently sensitive to brain damage, would produce an even more sensitive instrument. Since numerous studies (Anderson, 1951; Benton, 1953; Cohen, 1952; Collins, 1951; Diers & Brown, 1950; Graham & Kendall, 1946; Halstead, 1947; Heilbrun, in press; Hoedemaker & Murray, 1952; Ptacek & Young, 1954; Reynell, 1944; Ruesch & Moore, 1943) have demonstrated various tests of immediate memory to be differentially difficult for brain-damaged *S*s, the new procedure should provide evidence regarding the proposed hypothesis.

Method

Subjects

The 83 *S*s used in this study were patients at the University and Veterans Administration Hospitals, Iowa City. This number included 40 brain-damaged *S*s with confirmed cerebral pathology, obtained from neurological or neurosurgical services, and 43 physically ill normal *S*s who had no history of nor displayed any evidence of cerebral dysfunction.

Each diagnostic group was divided into two experimental subgroups to be described later, all groups being appropriately matched for

---

[1] This investigation was supported by a research grant (B-616) from the National Institute of Neurological Diseases and Blindness of the National Institutes of Health, United States Public Health Service.

Table 1

Age and Education Data for the Brain-Damaged and Physically Ill Control Groups

		Age		Education	
Group	*N*	Mean	*SD*	Mean	*SD*
Brain-damaged 1[a]	21	36.10	9.18	10.67	2.50
Physically ill 1[a]	23	37.96	8.32	10.26	3.11
Brain-damaged 2[b]	19	36.79	11.73	10.68	2.27
Physically ill 2[b]	20	34.50	10.17	9.85	2.56

[a] These groups were directly compared in the study.
[b] These groups were directly compared in the study.

age and education. The data for these age and education variables are presented in Table 1. None of the differences in means or variability approached statistical significance.

No *S*s over 50 years of age were included in the study so as to control for the detrimental effects of aging upon this type of intellectual performance. Also, any *S* who was not able to perform correctly on at least half the block design items administered under standard conditions was eliminated. This requirement excluded grossly impaired *S*s from consideration.

Task

Twenty block designs, collected from the two forms of the Wechsler-Bellevue Intelligence Scale and the Kohs designs or constructed by *E*, were combined into a set (Form A). Each design could be reproduced by using four Wechsler blocks. A parallel form of this task (Form B) was then constructed by color reversal, mirror reversal, or rotation of the designs. A pilot study using 20 college *S*s found these two tasks to be of equal difficulty when administered by an immediate memory procedure. Two designs on each form were used as demonstration items, leaving 18 test items to be scored on each. The items in Form A were placed in order of increasing difficulty as determined by the judgments of 15 experienced clinical psychologists. Each Form B item was automatically placed by the judged position of its matched item on Form A.

The standard block design task required the *S* to copy four block designs within a 60-sec. time limit with the design continuously in view for reference.

With the immediate memory block design procedure, a design was exposed for 10 sec. and then removed. An *S* was then given four blocks with which to construct the design from memory within the 60-sec. time limit.

Scoring

The scoring for block design performance was modeled after and similar to that used on the Wechsler-Bellevue. Four points were assigned if a design were constructed without errors within the 60-sec. time limit. Further, the following speed credits were given when a design was accurately reproduced: three points, 1–15 sec.; two points, 16–30 sec.; one point, 31–45 sec. Accordingly, a maximum score of 126 points was possible for the 18 items.

Either block design task was discontinued after *S* made errors on five consecutive trials.

Procedure

Approximately half of each diagnostic group was assigned to two order conditions. In one order condition, *S*s were given the immediate memory procedure first and the standard procedure second, while the remaining brain-damaged and control *S*s were given the procedures in the reverse order. In either case, the two measures were separated by interpolated psychological tests requiring about a half hour to administer. Within both order condition, *S*s were alternately given the memory procedure on Form A and Form B.

Results

The hypothesis of differential difficulty for brain-damaged *S*s on the memory procedure was tested by comparing the scores on the standard and memory tasks for each *S*. These performance difference scores were then averaged over *S*s in the brain-damaged and control groups. The critical test of the hypothesis was the difference between these average group difference scores.

Table 2 presents the performance comparisons for the two block design procedures under the two order conditions. It can be seen that the brain-damaged group showed a greater difference in scores for the two ad-

Table 2

Differences in Mean Scores for Brain-Damaged and Control *S*s for Two Block Design Procedures Under Two Order Conditions

Order Condition	Brain-Damaged			Control		
	Regular Procedure	Memory Procedure	Difference	Regular Procedure	Memory Procedure	Difference
Memory procedure first	98.62	41.34	57.28	109.56	65.65	43.91
Memory procedure second	93.26	46.84	46.42	101.95	68.55	33.40

ministration procedures, both when the memory test was given first and when it followed the standard test. The difference between the mean difference scores was significant both under the memory-first condition ($t = 2.34$ for 42 *df*, $p < .02$) and under the memory-second condition ($t = 2.24$ for 37 *df*, $p < .03$). These results clearly indicate that the immediate memory block design procedure is differentially difficult for brain-damaged *S*s as compared with control *S*s when performance on the standard block design procedure is used as a baseline.

Discussion

The hypothesis that the combination of two psychological processes, each independently sensitive to cerebral pathology, would provide a complex task of increased sensitivity to pathological brain states was supported by the results of this study. Thus the revision of the standard block design procedure, using rather simple four-block designs, by administering it as an immediate memory test provided a relatively more difficult task for brain-damaged *S*s than controls when compared with performances of these groups on the standard task. This relationship held whether the memory procedure was presented first and the standard procedure second or whether this order of presentation was reversed. Certainly one implication of these findings would be that further research direct itself towards evaluation of more complex processes as indicators of cerebral pathology.

The question could be raised as to how the block design memory task compares in discriminative efficiency to the standard task used in the present study and to a more difficult block design task using standard procedures.

An analysis of the overlap data for the memory task in the current study found that under the memory-first condition, the optimum cutting score provided correct classification for 14 of 21 brain-damaged *S*s and 18 of 23 control *S*s, an average efficiency of about 73%. Under the memory-second condition, an optimum cutting point on the memory task led to correct classification for 12 of 19 brain-damaged *S*s and 15 of 20 control *S*s, with an average efficiency of about 69%.

In comparison to these average discriminative efficiency figures of about 73% and 69% for the memory task, the standard task used in the present study maximally identified about 70% and 65% under the memory-first and memory-second conditions. These standard procedure figures are based upon correct classification of 12 out of 21 brain-damaged *S*s and 19 of 23 control *S*s (memory-first condition) and 16 of 19 brain-damaged *S*s and 9 of 20 controls (memory-second condition).

The results of a previous study by Heilbrun (1954) provide a basis for comparing the memory task with a more difficult standard block design task. In the earlier study, 100 *S*s (71 brain-damaged *S*s, 29 physically ill control *S*s), selected from the same hospitals as the *S*s in the present study, were given the block design subtest from the Wechsler-Bellevue under standard conditions as part of a larger test battery. It was found that this block design test provided better discrimination between brain-damaged *S*s and controls than any of the other 14 tests in the battery, accurately classifying an average of about 72% of the Ss in these two groups after selection of an optimum cutoff score.

The diagnostic comparison between the immediate memory block design procedure and the standard Wechsler block design test does not suggest that the former provides a more sensitive task for the detection of brain-damaged Ss. Further, the comparison between the memory procedure and the standard task utilized in the present study does not favor the memory procedure as a discriminative instrument in any clear-cut manner. However, the scoring system used in this study was chosen because it seemed to best allow for a direct comparison between the current standard procedure with block designs and a modified procedure. It seemed of interest to try at least one other scoring system to better judge whether the obtained levels of discriminative efficiency represented actual upper limits of the memory task for these samples of Ss or artificial limits imposed by the mode of scoring. It was decided to score by the number of individual blocks (i.e., there are four individual blocks per design) correctly placed within the time limit with no time bonus scores. This more precise scoring for accuracy of performance, again using an optimum cutting score, correctly classified 17 of 21 brain-damaged Ss and 17 of 23 control Ss under the memory-first condition, an average correct placement of about 78%. Under the memory-second condition, optimum cutting led to correct classification for 13 of 19 brain-damaged Ss and 16 of 20 controls, an average efficiency of about 74%. Using this more precise accuracy score, the maximum efficiency for the standard procedure was approximately 65% for the memory-first condition (10 of 21 brain-damaged Ss and 19 of 23 controls correctly identified) and about 62% for the memory-second condition (14 of 19 brain-damaged Ss and 10 of 20 controls correctly identified). These findings, though based on small samples, give more support to the notion that the present memory procedure compares favorably to either the standard procedure using simple four-block designs or the standardly administered and scored Wechsler task relative to the prediction of cerebral pathology.

It should be noted that using this more precise error scoring system, significant differences between brain-damaged and control Ss in their relative performances under the two administrative procedures were demonstrable. Under the memory-first condition the t value was 3.51 for 42 df ($p < .001$), while under the memory-second condition the t value was 2.64 for 37 df ($p < .01$). In both comparisons, the brain-damaged showed the greater relative impairment on the memory task.

The finding that discriminations between brain-damaged and control Ss was enhanced when bonus scores were *not* assigned for speed of performance (in addition to the more precise accuracy scoring) raised the interesting possibility that speed of performance did not show the differential decline when the memory block design procedure was compared with the standard procedure. Analysis of performance speed showed that brain-damaged Ss were consistently slower than controls but that the speed differences between these diagnostic groups *decreased* under the memory procedure, though the trend did not reach statistical significance. This finding stands in contrast to the finding that brain-damaged Ss show a reliable *increase* in impairment with the memory procedure when only performance accuracy is considered.

It should be emphasized that the findings of the present study do not warrant utilizing *this* particular memory task as an individual diagnostic measure. This study was primarily concerned with the effects of combining block design and immediate memory *procedures* upon the performance of Ss with cerebral damage and not with the construction of a diagnostic test. Considerably more research would be necessary before any such practical application would be warranted.

Summary

1. This study investigated the hypothesis that the combination of two psychological processes, each independently sensitive to the effects of cerebral damage, would provide a complex task of increased sensitivity to cerebral pathology.

2. Performances of Ss with confirmed cerebral pathology and matched physically ill Ss showing no history or current evidence of

cerebral damage were compared on two block design procedures—the standard copying procedure and an immediate memory procedure.

3. The prediction of differential difficulty on the memory task for brain-damaged *S*s was supported by the data.

4. Consideration of the overlap data in this study and the relevant data from another study suggested the relative diagnostic usefulness of a block design memory procedure when a more precise evaluation of performance accuracy is utilized.

5. Comparison of the results obtained when accuracy and speed of performance were analyzed separately suggested that brain-damaged *S*s show differential impairment in accuracy of performance on the memory task but do not show differential impairment in speed of performance.

Received August 4, 1958.

REFERENCES

Aita, J. A., Armitage, S. G., Reitan, R. M., & Rabinowitz, A. The use of certain psychological tests in the evaluation of brain injury. *J. genet. Psychol.*, 1947, **37**, 25–44.

Allen, R. M. The test performance of the brain injured. *J. clin. Psychol.*, 1947, **3**, 225–230.

Allen, R. M. The test performance of the brain diseased. *J. clin. Psychol.*, 1948, **4**, 281–284.

Anderson, A. L. The effect of laterality localization of focal brain lesions on the Wechsler-Bellevue subtests. *J. clin. Psychol.*, 1951, **7**, 149–153.

Benton, A. L. L'application clinique du test de rétention visuelle. Paris: Centre de Psychologie Appliquée, 1953.

Boyd, H. F. A provisional quantitative scoring with preliminary norms for the Goldstein-Scheerer Cube Test. *J. clin. Psychol.*, 1949, **5**, 148–153.

Cohen, J. Factors underlying Wechsler-Bellevue performance of three neuropsychiatric groups. *J. abnorm. soc. Psychol.*, 1952, **47**, 359–365.

Collins, A. Louise. Epileptic intelligence. *J. consult. Psychol.*, 1951, **15**, 392–399.

Diers, W. C., & Brown, C. C. Psychometric patterns associated with multiple sclerosis. *Arch. Neurol. Psychiat.*, 1950, **63**, 760–765.

Goldman, R., Greenblatt, M., & Coon, G. P. Use of the Wechsler-Bellevue scale in clinical psychiatry with particular reference to cases with brain damage. *J. nerv. ment. Dis.*, 1946, **104**, 144–179.

Graham, F. K., & Kendall, B. S. Performance of brain-damaged cases on a memory-for-design test. *J. abnorm. soc. Psychol.*, 1946, **41**, 303–314.

Grassi, J. R. The Fairfield Block Substitution Test for measuring intellectual impairment. *Psychiat. Quart.*, 1947, **21**, 474–489.

Grassi, J. R. *The Grassi Block Substitution Test for measuring organic brain pathology.* Springfield, Ill.: Charles C Thomas, 1953.

Greenblatt, M., Goldman, R., & Coon, G. P. Clinical implications of the Bellevue-Wechsler test (with particular reference to brain damage cases). *J. nerv. ment. Dis.*, 1946, **104**, 438–442.

Halstead, W. C. *Brain and intelligence.* Chicago: Univer. Chicago Press, 1947.

Hécaen, H., de Ajuriaguerra, J., & Massonet, J. Les troubles visuoconstructifs par lésion pariéto-occiptale droite. Rôle des perturbations vestibulaires. *Encéphale*, 1951, **40**, 122–179.

Heilbrun, A. B. The localization of cerebral lesions by the use of psychological tests. Unpublished doctoral dissertation, State Univer. of Iowa, 1954.

Heilbrun, A. B. The digit span test and the prediction of cerebral pathology. *Arch. Neurol. Psychiat.*, in press.

Hoedemaker, E., & Murray, M. Psychological tests in the diagnosis of organic brain disease. *Neurology*, 1952, **2**, 144–153.

Lidz, T., Gay, J. R., & Tietze, C. Intelligence in cerebral deficit states and schizophrenia measured by Kohs block test. *Arch. Neurol. Psychiat.*, 1942, **48**, 568–582.

McFie, J., & Piercy, M. F. Intellectual impairment with localized cerebral lesions. *Brain*, 1952, **75**, 293–311.

McFie, J., Piercy, M. F., & Zangwill, O. L. Visual-spatial agnosia associated with lesions of the right cerebral hemisphere. *Brain*, 1950, **73**, 167–190.

Milner, Brenda. Intellectual effects of temporal-lobe damage in man. Unpublished doctoral dissertation, McGill Univer., 1952.

Paterson, A., & Zangwill, O. L. Disorders of visual space perception associated with lesions of the right cerebral hemisphere. *Brain*, 1944, **67**, 331–358.

Ptacek, J. E., & Young, Florence. Comparison of the Grassi Block Substitution Test with the Wechsler-Bellevue in the diagnosis of organic brain damage. *J. clin. Psychol.*, 1954, **10**, 375–381.

Reynell, W. R. A psychometric method of determining intellectual loss following head injury. *J. ment. Sci.*, 1944, **90**, 710–719.

Ruesch, J., & Moore, B. E. Measurement of intellectual functions in the acute stage of head injury. *Arch. Neurol. Psychiat.*, 1943, **50**, 165–170.

Shapiro, M. B. Experimental studies of a perceptual anomaly. I. Initial experiments. *J. ment. Sci.*, 1951, **97**, 90–110.

Shapiro, M. B. Experimental studies of a perceptual anomaly. II. Confirmatory and explanatory experiments. *J. ment. Sci.*, 1952, **98**, 605–617.

Shapiro, M. B. Experimental studies of a perceptual anomaly. III. The testing of an explanatory theory. *J. ment. Sci.*, 1953, **99**, 394–409.

Wechsler, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

PERSONALITY AND BACKGROUND FACTORS IN JUVENILE DELINQUENCY AS INFERRED FROM QUESTIONNAIRE RESPONSES[1]

DONALD R. PETERSON, HERBERT C. QUAY,[2] AND GORDON R. CAMERON
University of Illinois

Most investigations of personality factors in crime and delinquency have begun with a legally defined sample of offenders, proceeded with various comparisons between that group and a more or less carefully matched group of nonoffenders, and ended with ambiguous results. After reviewing 113 such comparisons, Scheussler and Cressey (1950) stated that "The doubtful validity of many of the obtained differences, as well as the lack of consistency in the combined results, makes it impossible to conclude from these data that criminality and personality elements are associated" (p. 476). Like all negative findings, however, these can be interpreted in at least two ways, viz., as evidence for essential identity between offenders and nonoffenders in respect to personality, or as consequences of methodological failure. We believe the latter interpretation to be more plausible than the first, and submit that the most glaring defects in many previous investigations lie in the gross behavioral heterogeneity of legal offenders and inadequacies in the instruments used to examine them. The first condition is likely to lead to serious attenuation of all observed relationships, and to eventuate in a most unimpressive array of low correlations and small differences. The second condition is likely to obscure true relationships beyond all recognition.

As correctives, we propose that greater care be exercised in defining and measuring certain personality traits of empirically demonstrated importance in the predisposition to delinquent activity, and that subsequent research be concentrated upon the origins and consequences of these tendencies, rather than in the direct but premature search for the causes and outcomes of delinquency itself. The response-based inference of constructs which relate to delinquency, the development of progressively more adequate measures of these constructs, and the patient delay of the search for causal antecedents until such steps are taken, is slower and more restrictive than most of the earlier approaches, but it promises a degree of explanatory depth and precision which the older methods can never afford.

The present study is an attempt to define such constructs through the factor analysis of two sets of questionnaire items previously shown to differentiate between delinquents and nondelinquents (Gough & Peterson, 1954; Quay & Peterson, 1958). The initial mode of item selection has guaranteed that these response tendencies have *something* to do with delinquency and, in fact, one of the item sets, comprising the Socialization Scale of the California Psychological Inventory, has displayed a degree of construct validity truly unusual among personality tests (Gough, 1957; Gough & Peterson, 1954; Peterson, Quay, & Anderson, 1959). Inspection of the items, however, reveals a diversity of content and meaning nearly as great as for delinquency itself, and this analysis was undertaken in the belief that factor analysis could lead to the eduction of a set of reasonably unitary, independent, possibly meaningful constructs that would offer more powerful hypotheses for further research than is provided by the original more heterogeneous measures.

---

[1] Parts of this research were supported by a grant from the Institute for Research and Training in the Social Sciences, Vanderbilt University.

[2] Now at Vanderbilt University.

Subjects and Procedure

In an effort to maximize variance along dimensions of primary concern, both delinquents ($N = 116$) and nondelinquents ($N = 115$) were included in the sample. To minimize variance along at least one irrelevant dimension, the analysis was restricted to data from white subjects (*S*s). Delinquents and nondelinquents had been matched in respect to age and place of residence for the purposes of another study (Peterson et al., 1959), the report of which contains more information about the selection of *S*s than this one can.

All *S*s were given a combined form of two previously developed "delinquency" scales (Gough & Peterson, 1954; Quay & Peterson, 1958), phi coefficients were obtained for all item pairs, and 15 factors extracted by the complete centroid method. Judgments based on Tucker's function and contributions to total variance led to the exclusion of 10 of the factors. The remaining five provide a highly efficient, if not exhaustively sufficient, accommodation of the relationships in the phi matrix.

Rotations were pursued in two ways. First, an orthogonal analytic solution was reached by means of the quartimax routine for the electronic computer (Neuhaus & Wrigley, 1954). This yielded an acceptable but not strikingly impressive solution, in terms of the usual criteria of simple structure. All hyperplanes were significant at the .05 level by Bargmann's (1954) test, but visual inspection of the plots suggested that the solution might be improved somewhat by further shifts. The centroid factors were then rotated to an oblique analytic solution by means of the oblimax routine for the computer, and this was followed by a series of visually directed rotations until a virtually unimprovable solution had been found. All hyperplanes defined by these latter operations were significant far beyond the .001 level. Intercorrelation of loadings, however, demonstrated that the two independent solutions were very closely similar (*r*'s for analogous factors ranged from .89 to .99), and the simplicity and objectivity of the orthogonal quartimax solution led to its choice for presentation here. The factors given below are utterly free of contamination by interpretative bias, offer the simplicity of an orthogonal space, and

Table 1

Factor 1: Psychopathy

No.	Loading	
18	64	The only way to settle anything is to lick the guy.
31	62	Winning a fight is more fun than anything.
17	62	The people that run things are usually against me.
12	61	Cops usually treat you dirty.
27	60	If you don't have enough to live on, it's OK to steal.
24	54	A lot of times it's fun to be in jail.
35	53	The only way to make big money is to steal it.
47	52	A person is better off if he doesn't trust anyone.
34	51	If the cops don't like you, they will get you for anything.
46	48	Life usually hands me a pretty raw deal.
22	46	Cops and judges will tell you one thing and do another.
21	43	A guy like me hits first and asks questions later.
6	43	It's dumb to trust older people.
51	43	I would do almost anything on a dare.
33	43	If somebody does something to me, I always get them back.
30	42	Most brothers and sisters are more trouble than they are worth.
32	40	I don't mind lying if I am in bad trouble.
70	40	I go out of my way to meet trouble rather than try to escape it.
40	39	I do what I want to, whether anybody likes it or not.
14	−38	I would rather be at home when things go wrong.
20	35	I got (or used to get) into a lot of fights in school.
98	34	I never cared much for school.
84	33	I have never done any heavy drinking.
19	33	I have run away from home because my folks treated me bad.
10	32	I'm really too tough a guy to get along with most kids.

Table 2

Factor 2: Neuroticism

No.	Loading	
58	68	I often feel that I am not getting anywhere in life.
49	56	Sometimes I used to feel that I would like to leave home.
67	56	I seem to do things that I regret more often than other people do.
83	55	I have often gone against my parent's wishes.
64	54	My parents often disapproved of my friends.
100	52	I sometimes wanted to run away from home.
94	52	I often feel as though I have done something wrong or wicked.
95	51	I don't think I'm quite as happy as others seem to be.
93	48	People often talk about me behind my back.
101	47	With things going as they are, it's pretty hard to keep up hope of amounting to something.
68	44	I would rather go without something than ask for a favor.
42	43	I sometimes feel that I made the wrong choice in my occupation.
63	42	I have very strong likes and dislikes.
66	40	I often act on the spur of the moment without stopping to think.
41	39	I get nervous when I have to ask someone for a job.
3	39	Sometimes I feel (or used to feel) that if I could just get away from home, everything would be all right.
22	38	Cops and judges will tell you one thing and do another.
75	37	In school I was sometimes sent to the principal for cutting up.
4	36	My folks usually blame bad company for the trouble I get into.
87	−35	Most of the time I feel happy.
69	34	I have more than my share of things to worry about.
81	33	It is hard for me to act natural when I am with new people.
11	32	It isn't their fault that most guys get into trouble.
7	32	My folks have sometimes been in trouble with the law.
28	32	When I was a little kid, I was always doing things my folks told me not to.
96	32	I used to steal sometimes when I was a youngster.
74	−30	I have never been in trouble with the law.

are, for all practical purposes, equivalent to an oblique solution which is unassailable on the grounds of simple structure.

Results

The combined questionnaire as it was actually administered, the centroid factor matrix, and the rotated factor matrix are reported elsewhere.[3] The factors will be defined here, in order of descending variance, by statement of the items which have loadings of .30 or more. A positive sign for any loading implies that the response "true" is associated with the positive pole of the factor; a negative sign indicates that a "false" response is associated with the positive pole of the factor. Item numbers refer to the variables as they appeared in the questionnaire as used herein and as deposited with the American Documentation Institute.

Tough, amoral, rebellious qualities are obviously implied by the variables given in Table 1, and these meanings, together with those of impulsiveness, a conspicuous distrust of legal and other authority, and an apparent freedom from family ties, led to our designation. Factor meaning seems very close to that of "Unsocialized Aggression," as identified in the work of Jenkins and his colleagues (Hewitt & Jenkins, 1946; Jenkins & Glickman, 1947), and strongly involves the same disregard for public opinion so prominent in the "Psychopathic Personality" dimension which emerged from Comrey's (1958) factor analysis of the Psychopathic Deviate (*Pd*) scale of the MMPI.

Tendencies toward impulsive action again appear in Factor 2, but the intrapsychic concomitants, and perhaps the dynamic bases, of the behavior are decidedly not psychopathic.

---

[3] The combined questionnaire, centroid factor matrix, and rotated factor matrix have been deposited with the American Documentation Institute. Order Document No. 5920, remitting $1.25 for microfilm or $1.25 for photocopies.

Table 3

Factor 3: Family Dissension

No.	Loading	
39	61	My mother and father argue a lot.
36	60	My step-father (or step-mother) treats me badly.
104	−49	My home life was always very pleasant.
103	49	I was often punished unfairly as a child.
99	−46	The members of my family were always very close to each other.
37	41	My folks yell at us kids a lot.
13	39	My mother and father have never really been friends of mine.
65	−39	My home life was always happy.
56	38	My home life as a child was less peaceful than those of most other people.
19	36	I have run away from home because my folks treated me bad.
29	36	I have lived in an orphans' home or a foster home at some time.

Remorse, tension, guilt, depression, discouragement—in short, neurotic responses—covary with antisocial activity. The factor bears strong resemblance to a factor labeled "Disturbed Delinquency" by Jenkins and Glickman (1947), and to the "Neuroticism" factor isolated by Comrey (1958).

This is clearly a family background factor. The items are so uniform in meaning that no further discussion is necessary.

Unlike the previous dimensions, Factor 4 is difficult to interpret. It was principally the pervasive sense of incompetence and failure which suggested "Inadequacy" as an interpretative label, and the general impression is indeed that of an inability to cope with the problems of a complex world. The meaning of the factor, however, is obscure and our interpretation tentative.

Table 4

Factor 4: Inadequacy

No.	Loading	
74	−43	I have never been in trouble with the law.
2	41	I am behind at least a year in school.
38	39	I'd quit school now if they would let me.
92	38	When I was going to school I played hooky quite often.
82	35	When something goes wrong I usually blame myself rather than the other fellow.
62	34	I hardly ever get excited or thrilled.
80	31	I enjoy work as much as play.
25	30	My folks move (or used to move) from place to place a lot.
44	30	I would have been more successful if people had given me a fair chance.

This factor is no easier to interpret than Factor 4. Only three of the items, however, attained very high loadings, and these all stand in some relationship to a history of conflict with school authority. The title we have tentatively attached implies such a history.

Table 5

Factor 5: Scholastic Maladjustment

No.	Loading	
92	48	When I was going to school I played hooky quite often.
75	47	In school I was sometimes sent to the principal for cutting up.
52	47	As a youngster in school I used to give the teachers lots of trouble.
76	−34	I keep out of trouble at all costs.
32	34	I don't mind lying if I am in bad trouble.
98	33	I never cared much for school.
96	33	I used to steal sometimes when I was a youngster.
57	−31	I think I am stricter about right and wrong than most people.
91	30	I have used alcohol excessively.
83	30	I have often gone against my parents' wishes.

Summary

Research on personality factors in delinquency has long been impeded by the gross behavioral heterogeneity of delinquents, and various inadequacies in the measures used to study them. A factor analysis of two questionnaire scales of demonstrated effectiveness in differentiating delinquents from nondelinquents was conducted in the belief that future research on the origins and consequences of mediating personality tendencies thus defined would lead to greater scientific progress than direct investigation of legally defined delinquency itself. Three personality dimensions and two background factors emerged. The first was characterized by a number of psychopathic qualities and was named accordingly. In the second factor, impulsive antisocial behavior covaried with expressions of regret, depression, and other negative affect. It was interpreted as a neurotic dimension. The third putative personality factor implied a general sense of incompetence and was regarded as an expression of inadequacy. Of the two background factors, one clearly related to family dissension, and the other seemed, much less clearly, to relate to a history of difficulty in school.

Received August 7, 1958.

REFERENCES

Bargmann, R. Signifikanzuntersuchungen der einfachen struktur in der faktorenanalyse. *Mitteilungsblatt f. math. Statist.*, 1954, **Nov.**, 1–24.

Comrey, A. L. A factor analysis of items on the MMPI Psychopathic Deviate Scale. *Educ. psychol. Measmt*, 1958, **18**, 91–98.

Gough, H. G. *Manual for the California Psychological Inventory*. Palo Alto: Consult. Psychologists Press, 1957.

Gough, H. G., & Peterson, D. R. The identification and measurement of predispositional factors in crime and delinquency. *J. consult. Psychol.*, 1954, **16**, 207–212.

Hewitt, L. E., & Jenkins, R. L. *Fundamental patterns of maladjustment: The dynamics of their origin*. Springfield, Ill.: Green, 1946.

Jenkins, R. L., & Glickman, Sylvia. Patterns of personality organization among delinquents. *Nerv. Child,* 1947, **6**, 329–339.

Neuhaus, J. O., & Wrigley, C. F. The quartimax method: An analytic approach to orthogonal simple structure. *Brit. J. statist. Psychol.*, 1954, **7**, 81–91.

Peterson, D. R., Quay, H. C., & Anderson, A. C. Extending the construct validity of a socialization scale. *J. consult. Psychol.*, 1959, **23**, 182.

Quay, H. C., & Peterson, D. R. A brief scale for juvenile delinquency. *J. clin. Psychol.*, 1958, **14**, 139–142.

Scheussler, K. E., & Cressey, D. R. Personality characteristics of criminals. *Amer. J. Sociol.*, 1950, **55**, 476–484.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

THE TWISTED PEAR AND THE PREDICTION OF BEHAVIOR [1, 2]

JEROME FISHER
Veterans Administration Hospital, San Francisco, California

"To say that nature should conform to a Gaussian distribution is asking too much. Who is there to tell nature what the statisticians would like?" This statement, made recently by Boring (1957) in an editorial context, seems an appropriate introduction to the question and the point of view presented in this paper. It is by no means a new point of view. Rather, it is a variation on the familiar theme that organisms, whose behaviors are variously studied for purposes of prediction, do not conform to the assumed mathematical conditions which are often assigned to them in statistical manipulations of data.

In substance, the paper raises questions not only about the appropriateness of our statistical assumptions as they concern prediction problems but, also, it ventures the proposition that, because of certain biologic and psychologic variants, organismic behaviors are predictable in only one segment of a predictor-criterion relationship.[3]

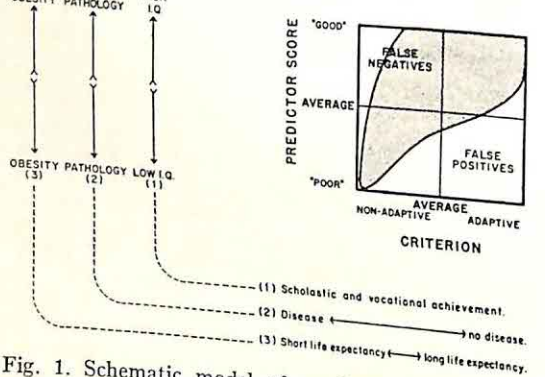


Fig. 1. Schematic model of prediction in behavior. Some examples: Predictor vs. Criterion.

Several years ago, in analyzing the results of a cross-validation study (Fisher, Gonda, & Little, 1955) involving the Rorschach, my colleagues and I noted a curious but consistent result. When the Rorschach yielded a score indicating the presence of brain disease, the agreement with independent criterion judgments of brain pathology was extraordinarily good (94%). A low score, however, was *not* accurate in predicting the absence of brain pathology. A Pearson validity coefficient was computed and found to be a respectable, significant, but humble .32.

Next, we subjected several standard neurologic procedures, such as the EEG, the lumbar puncture, etc., to essentially the same validation analysis (Fisher & Gonda, 1955). Except for a few minor variations, five neurologic diagnostic techniques gave the same results: High accuracy in predicting pathology from positive test findings, but like the Rorschach, poor accuracy in predicting the absence of pathology from negative diagnostic signs. The over-all validity coefficients ranged from .13 to .32.

When scattergrams were plotted of the relationship between predictor scores and the criterion, a nonlinear heteroscedastic configuration was revealed which looked like a twisted pear; it is approximated in Fig. 1. This finding raised the question: Is the twisted pear unique to these data or is it a general pattern, characteristic of prediction problems?

---

[1] Based on a paper read at the XV International Congress of Psychology, Brussels, July, 1957.

[2] For many helpful sugestions and for their encouragement the author is especially grateful to Robert C. Tryon and Harrison Gough.

[3] Cronbach and Gleser (1957) have recently presented a similar argument to the one advanced in this paper, based on the mathematics and logic of decision theory, for considering the problem of differential validities in psychological testing.

The diagram will call to mind many familiar observations and findings. When scattergrams from several sources are examined, where more or less known criteria are plotted against more or less standard predictors, e.g., objective tests, a linear relationship with relatively little array variance will be observable at one end of the plot. The relationship, however, becomes increasingly nonlinear and increasingly variable as it approaches the middle and upper extreme. The terms "adaptive" and "nonadaptive" are used in Fig. 1 to suggest an implicit but functionalistic conception of the corresponding relationships between predictor "good" and "poor" scores and the criterion behavior of organisms. In other words, the predictor becomes *decreasingly* predictive of the criterion as the scores obtained increase from the "poor" to "good" extremes of the predictor. To illustrate this predictor-criterion differential, the findings of studies of intelligence, learning, personality, and pathology will be presented and discussed.

According to Fig. 1, the predictive efficiency of the IQ at the "poor" extreme should be considerably better than for average and "good" (superior) IQs. Specifically, with IQs below, say 50, it is highly predictable that the individual will require custodial or complete protective care, and that he will not acquire any scholastic skills. With IQs of about 50–70, there appears to be moderately good predictability that the individual will require special training and guidance, particularly at work and at school. With IQs of approximately 70–85, and, to a greater degree between 85–110, however, vocational and scholastic limitations vs. successes become increasingly difficult to predict.

Table 1 summarizes the results of several longitudinal studies of the vocational and educational achievements of male mental defectives, their controls, and, in addition, the vocational and scholastic attainments of Terman's gifted men (Terman & Oden, 1947). The latter included some 730 subjects whose childhood IQs were 140 or higher and who were classified into two groups by raters on the basis of vocational achievement 18 years later. The A group (mean childhood IQ = 155) was the most successful and the C group (mean IQ = 150) the least successful. The C group of 150 men, therefore, was, by definition and rating procedures, an underachieving group. Table 1 reveals that as a group, Terman's gifted men, 25 years of age and older in 1940, were superior occupationally to the general male population in the U. S. Further analysis of the data (not reported in Table 1), however, discloses that 28% of the C group fell at or below the median occupational level of the employed males in the U. S., as well as of those in California in 1940.[4, 5] Terman and Oden (1947) note that:

> On the average, those of highest IQ accomplish more and are equally well-adjusted, but one cannot anywhere draw an arbitrary IQ line that will set off potential genius from relative mediocrity. Some of our subjects who have achieved most notably did not, either in childhood or in adult life, rate above the average of the total group in tested intelligence.

With regard to the lower end of the IQ range, Table 1 also gives the occupational achievements of two follow-up studies of mentally defective groups. The data reported are in the form of average percentages for the combined groups. Compared to the gifted and control groups, the findings suggest a stronger, less variable relationship between mental deficiency, as determined mainly by IQ, and occupational achievement. None of the mental defective groups achieve I and II occupational categories; 23% are in III and IV, and 77% in V and VI. For the control samples, whose mean IQs are within the average range, the proportion of cases fall in the several categories as shown. Approxi-

---

[4] See p. 361 ff. in Terman and Oden (1947).

[5] The unemployed and incapacitated ($N = 16$) in Terman's gifted group were not included in the numbers and percentages given in Table 1; in addition, those who were classified as students were also omitted. A mitigating consideration is the postdepression economy of the period and the relative employment "youthfulness" of the entire gifted group. Many were still in transit occupationally.

The same conditions, however, obtain for the mentally defective and their control groups, because all of the studies cited in Table 1 were based upon the socioeconomic conditions of the decade 1930–1940. In personal communication with the author, Melita H. Oden reported that since 1940, follow-up studies of the gifted group have shown intragroup shifts on the occupational scale; some C men have moved up and others have moved down on the scale; some A men have taken positions in lower (III-VI) occupations.

Table 1

Comparison of Educational and Occupational Achievements of Groups of Men Whose IQs Were Obtained in Childhood

Groups	Mean IQ at Time of Test	Mean Age at Time of Test	Follow-up in Years	Occupational Achievement				Educational Achievement				
					Minnesota Occupational Scale[a]							
				N	VI & V	IV & III	II & I	*N*	8th Grade and Lower	9th–12th	Some College	College Graduation
Terman's gifted (1940)												
A (*N* = 96)	155.0	9.7	18	149[b]	10.3%	24.3%	65.3%	150	0.0%	14.7%	18.3%	66.9%
B				436				481				
C (*N* = 92)	150.0			139				150				
Baller's (1935) control	107.8	8–9	21	113	43.5	40.6	15.9	124	36.0	35.3	15.5	13.1
Fairbank's (1931) control	90.0	12 (?)	17	39				25				
Baller's (1935) mental defectives	60.5	8–9	21	67	77.3	22.7	0.0	113	93.5	6.5	0.0	0.0
Fairbank's (1931) mental defectives	68.0	12	17	69				43				
U. S. Census (1940)[c]					44.7	41.6	13.8		60.4	29.5	5.5	4.6

[a] The Minnesota Occupational Scale was used by Terman for the classification of his gifted men. Because the other follow-up studies cited here were conducted and reported in the same decade (1930–1940), the Minnesota Occupational Scale was adapted and applied to them also. Roe's (1956) more recent two-way classification of occupations gives similar results. Roughly, the classifications I & II refer to professional and managerial; III & IV—skilled labor and agricultural; V & VI—semiskilled and laborer.

[b] The *N*s of 149 and 139, respectively, as well as the *N*s and percentages of the other studies cited, refer to gainfully employed men only, who were available geographically at the time of the follow-up study.

[c] U. S. census (1951) educational and occupational data for men are included for purposes of general comparison.

mately two-thirds of the gifted group attained superior vocational levels (I and II). The presence of 10 and 24% of the gifted in the lower occupational Categories V and VI, and III and IV, respectively, however, suggests again the greater variability and, hence, the potential predictive "error" associated with "good" IQ scores.

Prediction of scholastic achievement by means of the IQ was studied by review of the same investigations. Table 1 also presents the follow-up findings of the same group of men whose IQs were obtained in childhood. The data are summarized in scatterplot form via averages to make them comparable to Wolfle's (1951) categories of educational achievement.

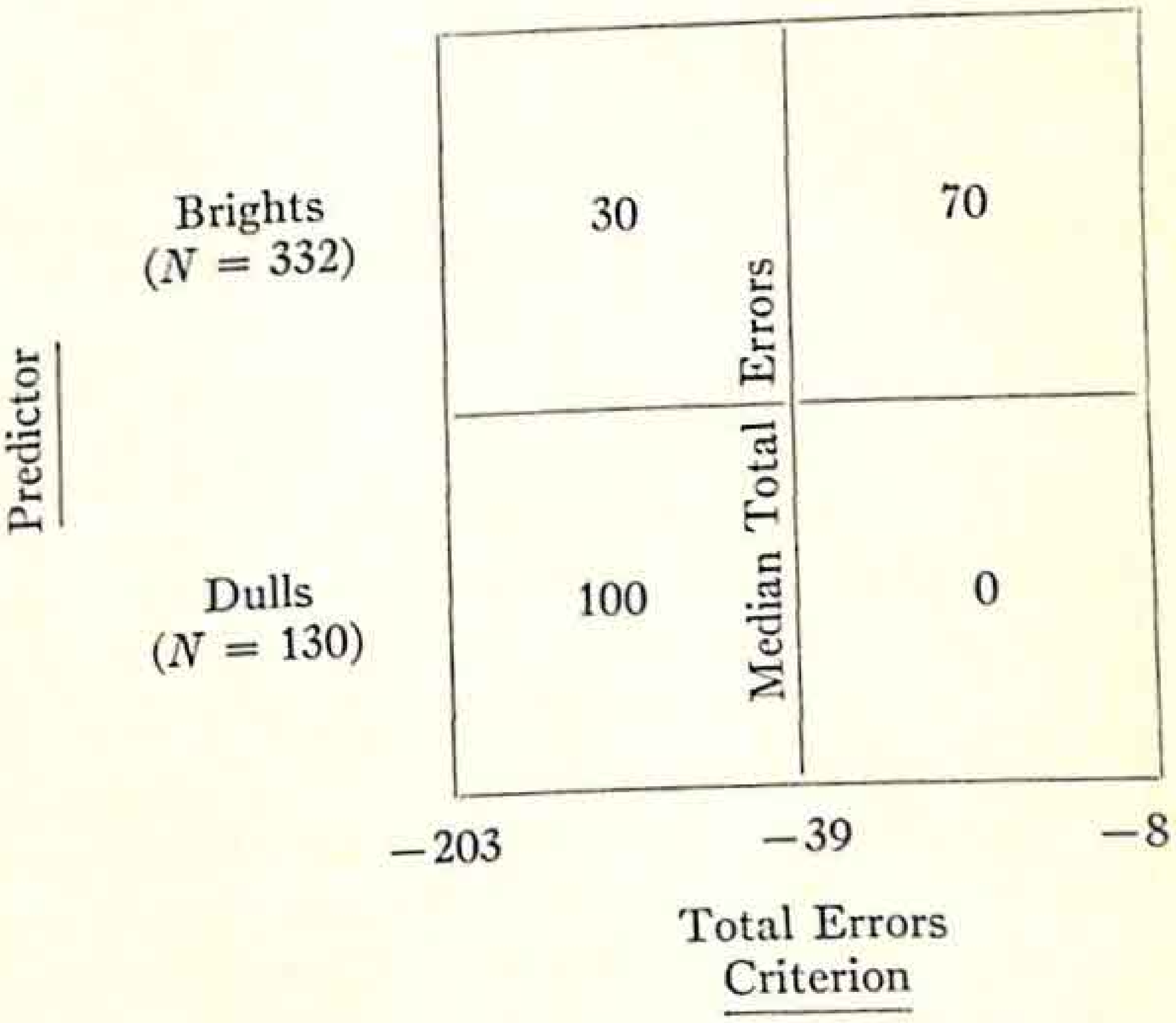
It appears that the IQ has considerable power in predicting scholastic achievement, particularly in the tails of the distribution of IQs. Compared to the controls and the gifted, however, the lower end of the IQ range (the mental defectives), reveals greater certainty of prediction for the criterion scholastic achievement.

At the upper extreme, the gifted's superiority educationally is self-evident (67% graduated from college). According to Terman and Oden, none of the gifted failed to complete grade school. For the middle and upper range of the predictor, however, the computed percentages, placed in their respective cells, suggest the characteristic variability observed before and again reveal the twisted pear shape.

What about learning, personality, and the "twisted pear" phenomenon? In his genetic research of dull-bright rats and maze-learning ability, Tryon (1940) found that while the poor maze-learning of the dull rats could be predicted very well, the bright rats varied throughout a wide range of learning scores (errors). Table 2 presents a fourfold table which was obtained by plotting around the median value of the total errors of the seventh generation ($N = 153$), the total errors made by their dull and bright offspring of the 15th to 18th generation. The reason for using the seventh generation as the source for the median value comes from Tryon's observation, "There appears to be a law of diminishing returns, for after the F_7 negligible effects of selective breeding are noted." None of the descendants of the dulls fell in the bright category, whereas 30% of the descendants of the bright ancestors performed in the dull category. The results, in short, suggest that the relation between predictor ancestors and criterion descendants, involving the learning of a complex set of highly integrated acts, is that of the twisted pear.

Table 2

Percentage of Bright and Dull Strains 15th–18th Generations, Who Fall Above and Below the Median Total Errors of the 7th Generation ($N = 153$) (From Tryon, 1940)



The work of the California-Berkeley group on authoritarianism suggests a similar pattern of differential predictability in personality assessment (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). While high scorers (authoritarians) and low scorers (nonauthoritarians) "emerge as a result of statistical analysis . . . they consist in accumulations of symptoms frequently found together but they leave plenty of room for variations of specific features. Furthermore, various distinct subtypes are found within each of the two major portions." Analysis of the prejudiced subjects, however, revealed that they are "on the whole more alike as a group than are the unprejudiced. The latter include a great variety of personalities . . ." (pp. 971–972). It seems, therefore, that a high score on the ethnocentrism scale is more predictive of authoritarianism than are low scores of its absence.

Leaving psychology for a moment, an analogue seems to exist in the actuarial determinations of the life insurance field (see Fig. 1). For example, having reviewed their experience tables and having ascertained the extent of their errors of prediction, insurance companies confidently relate obesity to a short span of life and set their premium rates accordingly. As weight approaches nonobesity, however, there appears to be increasing variability in predicting life expectancy. It is, of course, true that no one dies of obesity and the nonobese may die of many other causes; yet as a predictor variable of the probability of life expectancy, insurance companies respect the highly significant relationship found at the "poor"-nonadaptive extreme, i.e., between obesity and short life expectancy.

In their daily practice, my hospital medical colleagues confirm the curvilinear relationship between their tests and their criteria of pathology; at least, they acknowledge implicitly the application of differential predictability. For example, it seems that medical specialists almost always regard a positive test finding ("poor" predictor score) with considerable respect. The reason for this is that their diagnostic techniques are accurate at least 80% of the time when they yield a positive finding, i.e., the false positive rate is low. A negative or even a borderline finding, ("good" to average score) however, is invariably disregarded, *if*, in the physician's opinion, this finding is at variance with the patient's history, his presenting complaints, his symptoms and other diagnostic data. With a negative ("good" score) finding, therefore, the physician functions with a clinical relativism and a clinical tolerance of variability and errors of prediction.

When we apply our correlation statistics to measure the magnitude of the relationship between a criterion and a predictor, the coefficient thus obtained gives an average, over-all statement of differential predictions of the test. The result, being a statement of the weighted means of the selection ratios of the predictor classes $Y_1, Y_2, \ldots Y_n$ on the criterion X, is to attenuate the validity coefficient. If the product-moment r is computed without drawing the scatterplot, the plot may or may not be heteroscedastic, and the analyst cannot tell whether there are differential predictions of the various Y classes.

Guilford (1956), among others, has argued for the importance of scattergram inspection. With respect to organismic behavior, therefore, the case of linear association and equal variance obtains and applies partially, but not throughout a predictor-criterion relationship. The case of curvilinear association and unequal variance, however, also applies and, therefore, both cases are relevant and deserve our differential appreciation.

It has been said many times before and in many different ways that the behavior of the adaptive, functioning organism is ordinarily highly complex. I think we mean by this a capacity for variable, substitutive, compensatory behaviors. Under the disruptive conditions of pathology or stress, for example, it is as if special homeostatic mechanisms provide a biologic and/or psychologic *smoke screen* of adaptiveness. These on-going, restorative processes may be largely responsible for the high rate of "normal" responses or false negatives. It is perhaps no accident that many of our most useful diagnostic methods have been devised to assess pathology *in status;* i.e., when the camouflage is no longer impenetrable, or when adaptiveness has been or is curtailed beyond a certain point. I am referring here, in particular, to Binet's original mental test—an extraordinarily fine screening device for mental deficiency.[6] It is precisely in the nonadaptive segment, then, that our predictors seem to prove themselves by virtue of their high rate of predictive accuracy. In this connection, it should be pointed out that they are not entirely without predictive power in the adaptive segment either, albeit negative errors lessen their accuracy. It seems, therefore, that when predictor measures are extrapolated beyond that "certain point," i.e., from "poor" to "good," the multivariable complexities of behavior multiply and interfere with our predictive accuracy. Hence, it appears that whatever validity a measure may possess for predicting *adaptive* behavior, it is likely to be more accurate in predicting nonadaptive behavior.

---

[6] Recently, Jones (1958) advanced the concept of polarity to explain the one-ended conceptual clarity of most psychologic scales. He proposes a method for testing interitem homogeneity at either or both ends of a continuum to determine the degree of scale polarity.

Perhaps the hidden determinant in all this is the criterion dimension itself extending from the nonadaptive to the adaptive. It, too, appears to have a partial range of certainty and "truth," probably because more is known about the defining criterion points of nonadaptiveness than about those of adaptive behavior. Our test construction procedures provide us with reasonably good techniques and measures of criterion nonadaptive behavior. If, however, the twisted pear phenomenon possesses the degree of generality suggested by the data reviewed, then, there is reason to question the assumptions for extrapolating from the nonadaptive to the adaptive extremes of predictor-criterion relationships.

Is it possible that the observed *partial* curvilinearity and differential predictive variance are merely artifacts of sampling (of our tests) or of our criterion determinations? Is the twisted pear solely a "clinical" phenomenon, or is it a more general characteristic of the prediction problem? Perhaps these questions are unanswerable at the present time or perhaps we are dealing with a reality of organismic behavior, one of the existentional dilemmas, to borrow a term from Erich Fromm, for the behavioral sciences. It is as if, on the one hand, there is general acceptance of the dynamic nature of organismic adaptation and change, and the variability thereby induced, while, on the other hand, there is the coexisting pursuit of immutable validity coefficients as an attainable goal in the business of understanding and predicting behavior.

Received August 11, 1958.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

BALLER, W. R. A study of the present social status of a group of adults, who, when they were in elementary schools, were classified as mentally deficient. *Genet. psychol. Monogr.*, 1936, **18**, 165–244.

BORING, E. G. CP speaks. (Editorial.) *Contemp. Psychol.*, 1957, **2**, 261.

CRONBACH, L. J., & GLESER, GOLDINE C. *Psychological tests and personnel decisions*. Urbana: Univer. Illinois Press, 1957.

FAIRBANK, RUTH E. The subnormal child—seventeen years after. *Ment. Hyg., NY,* 1933, **17**, 177–208.

FISHER, J., & GONDA, T. A. Neurologic techniques and Rorschach test in detecting brain pathology: A study of comparative validities. *AMA Arch. Neurol. Psychiat.*, 1955, **74**, 117–124.

FISHER, J., GONDA, T. A., & LITTLE, K. B. The Rorschach and central nervous system pathology: A cross-validation study. *Amer. J. Psychiat.*, 1955, **3**, 487–492.

GUILFORD, J. P. *Fundamental statistics in psychology and education*. New York: McGraw-Hill, 1956.

JONES, M. B. The polarity of psychological tests. *J. consult. Psychol.*, 1958, **22**, 25–29.

ROE, ANNE. *The psychology of occupations*. New York: Wiley, 1956.

TERMAN, L. M., & ODEN, MELITA H. *Genetic studies of genius*. Vol. 4. *The gifted child grows up*. Stanford: Stanford Univer. Press, 1947.

TRYON, R. C. Genetic differences in maze-learning ability in rats. *Yearb. nat. Soc. Stud. Educ.*, 1940, **39**, Part I.

WOLFLE, D. Intellectual resources. *Scient. Amer.*, 1951, **185** (3), 42–46.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

FACTOR ANALYSES OF RORSCHACH SCORING CATEGORIES AND FIRST RESPONSE TIMES IN NORMALS[1]

JULIUS WISHNER
University of Pennsylvania

The Scoring Categories

The present studies were undertaken and completed prior to the appearance of the factor analyses by Williams and Lawrence (1953, 1954) and Consalvi and Canter (1957). The study by Borgatta and Eschenbach (1955) in particular raises serious questions as to the predictability of overt behavior from the Rorschach test, as does the experiment of Holtzman and Sells (1954). Nevertheless, the Rorschach test continues to be used in the clinical setting, and there is continued interest in research results.

Opportunity to check on Wittenborn's (1950a) results of his factor analysis of the Rorschach scoring categories, taking advantage of suggestions stemming from Cronbach's (1949) critique of statistical procedures utilized by Rorschach researchers, was provided by the Rorschach data collected by Beck, Rabin, Thiesen, Molish, and Thetford (1950) on 157 subjects from a stratified normal population. These data also provided the opportunity to determine if factors of verbal productivity and psychological efficiency suggested by previous investigations (Wishner: 1948, 1953, 1955) were identifiable. Verbal productivity factors in the Rorschach have also been reported by Sen (1950), Fiske and Baughman (1953), Borgatta and Eschenbach (1955), and Wittenborn (1950a).

From the point of view of populations studied, Wittenborn's Yale undergraduates are the only nonpsychiatric subjects whose Rorschachs have been factor-analyzed. The present study deals with a nonpsychiatric population, less selected than Yale undergraduates, representing four general occupational groups: executives and junior executives, skilled workers, semiskilled, and unskilled as reported elsewhere (Beck et al., 1950). While the population studied by Eschenbach and Borgatta (1955) was also ostensibly normal, it did include military personnel who had been incarcerated for breaches of discipline; in addition, their population, in general, can be presumed to be as selective as the Air Force is in selecting its personnel.

From the standpoint of scoring and computational procedures, the present study is most like those of Eschenbach and Borgatta (1955). They dealt with the general problem of the dependence of the absolute number of responses in any particular scoring category on the total number of responses (R) by means of partial correlation. In this study, the scores were converted into percentages. Since many of the distributions were highly skewed, scattergrams were constructed of the relationship between the most highly skewed variables, and tests of linearity were made. Significant nonlinearity was found in one case, and this is discussed below.

Computational Procedures, Results, and Interpretation of Factors

The population is described in detail by Beck et al. (1950). There were 71 males and

---

[1] The original data for this study were collected under the supervision of Samuel J. Beck at the Michael Reese Hospital in Chicago. The author is indebted to S. J. Beck and the administrators of the Department of Psychiatry of the Michael Reese Hospital for making the raw data available. The major part of the statistical work was performed by Lillian Berg. The author is pleased to express his deep appreciation to the Committee on the Advancement of Research of the University of Pennsylvania for its financial support of this project. Appreciation is also due Malcolm G. Preston and Howard Maher for helpful suggestions and consultation.

Table 1

Rorschach Scoring Categories Analyzed

Variable No.	
1.	R
2.	$W\%$
3.	$D\%$
4.	$Dd\%$
5.	$M\%$
6.	$(C + CF)\%$
7.	$(C + CF + FC)\%$
8.	$FC\%$
9.	$Sum\ C/R$
10.	$(Y + YF + FY)\%$
11.	$(V + VF + FV)\%$
12.	$(Y + YF + FY + V + VF + FV)\%$
13.	$F\%$
14.	$F{+}\%$
15.	$H\%$
16.	$A\%$
17.	N/R (N = The number of content categories utilized by the subject)
18.	Z/R
19.	$P\%$
20.	T/R (Time per response)

86 females; the age range was 17 to 69, with a mean of 30.5 and an *SD* of 8.55. There were 39 executives and junior executives, 47 skilled workers, 44 semiskilled, and 24 unskilled. The subjects (*S*s) were administered the Rorschach by two different examiners. Fiske and Baughman (1953) report that there were no systematic differences between the examiners in total number of responses elicited, and that each tested the same proportion of the four vocational groups.

The scoring categories chosen for analysis are shown in Table 1. It will be seen that, for the most part, these are the standard Rorschach scoring categories as used by Beck. The C, Y, and V scores were broken up as they are in order to test again Wittenborn's findings that FC has a factorial composition different from C and CF, to test the more general implication of his findings concerning the dominance of F, and to test the assertion implied by the scoring system that Y and V represent different psychological qualities.

The product-moment intercorrelations of these scores were computed and are shown in Table 2. Blakeman's tests of linearity were applied to 12 comparisons involving the most skewed distributions, and yielded significant nonlinearity only between $H\%$ and $P\%$. However, since r of $H\%$ and $P\%$ is $-.04$, whereas η is only .33, and since inspection of scattergrams indicated no obvious curvilinearity in the other comparisons, it did not seem necessary to resort to transformations.

These intercorrelations were then factor-analyzed by the centroid method, and the results are shown in Table 3. As in most factor analyses of the Rorschach test, four factors emerge.

Attempts to rotate these factors according to Thurstone's criteria of simple structure did not meet with success. Plots of the factors against each other revealed that in most cases there was an approximately equal distribution of the tests in the four quadrants. Thus, there seemed to be no objective bases for preferring one rotation over another.

It was decided, therefore, to make direct tests of the hypotheses underlying this study, as well as tests of general Rorschach hypotheses. These hypotheses were:

1. That there is a factor of psychological efficiency;
2. That the factorial composition of FC is different from that of C and CF;
3. That there is a factor of verbal productivity.

The fourth hypothesis, added after the first three had been tested, was:

4. That there is a factor of compulsivity.

The procedure utilized in testing these hypotheses consisted of setting down the scores which are supposed to reflect the factor involved, choosing the central score of this complex, rotating the factors so that the loading on this central score is maximized, noting the loadings on the other scores, and then estimating the degree of agreement with the original hypothesis. For example, in order to test the hypothesis that a factor of psychological efficiency is identifiable in the Rorschach, the factors were rotated through $F\%$, since in a previous study it was found that $F\%$ showed the highest coefficients of correlation with general alertness (Wishner, 1953). It was hypothesized that the factor would have loadings in $F\%$, T/R, and $A\%$, as well as significant loadings of the opposite sign in $M\%$,

Table 2

Intercorrelations of Rorschach Scoring Categories

Variable No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2.	-.332																		
3.	.120	-.811																	
4.	.409	-.609	-.043																
5.	.166	.012	.029	-.027															
6.	-.004	.434	-.247	-.396	.046														
7.	-.007	.467	-.257	-.426	.023	.885													
8.	-.030	.253	-.128	-.256	-.018	.167	.599												
9.	-.010	.459	-.253	-.420	.044	.964	.951	.370											
10.	.122	.132	-.132	-.014	-.011	.194	.289	.293	.256										
11.	.177	.068	-.106	.066	.132	.155	.087	-.091	.144	.108									
12.	.194	.136	-.169	.031	.076	.237	.263	.161	.272	.801	.685								
13.	-.192	-.288	.162	.220	-.551	-.606	-.679	-.429	-.661	-.539	.087	-.650							
14.	-.251	.159	0	-.298	.071	.108	.142	.133	.145	.151	.185	.224	-.224						
15.	.208	-.226	.081	.305	.746	-.241	-.269	-.170	-.256	-.067	-.002	-.059	-.183	.057					
16.	-.424	-.065	.155	-.150	-.344	-.241	-.207	-.010	-.232	-.246	-.311	-.367	.467	.081	-.270				
17.	-.345	.357	-.240	-.263	-.186	.331	.307	.101	.326	.227	.039	.196	-.179	.064	-.334	-.298			
18.	-.217	.787	-.581	-.473	.254	.446	.496	.298	.469	.138	.157	.196	-.477	.232	-.048	-.221	.363		
19.	-.515	.206	-.011	-.342	-.008	-.127	-.093	.048	-.111	-.166	-.288	-.294	.198	.449	.016	.386	.083	.115	
20.	-.470	.117	-.110	-.035	-.104	-.255	-.290	-.186	-.283	-.174	-.168	-.231	.337	-.022	-.053	.449	-.018	-.054	.254

Table 3

Unrotated Factor Loadings—Scoring Categories

Variable No.	I	II	III	IV	h^2
1.	−.2229	−.5774	−.3565	.3627	.6417
2.	.5201	.5572	.1774	.3005	.7027
3.	−.4353	−.2430	−.1391	−.2901	.3520
4.	−.5314	−.4777	−.1359	−.1362	.5476
5.	.2758	−.4685	.4091	.4561	.6709
6.	.7080	.2013	−.4029	.2011	.7446
7.	.8132	.2439	−.4939	.2350	1.0199
8.	.4288	.1321	−.2274	.0509	.2556
9.	.7896	.2250	−.4624	.2092	.9317
10.	.5442	−.3301	−.2747	−.3436	.5986
11.	.4552	−.3933	.2264	−.2959	.5007
12.	.6298	−.5345	−.1481	−.4301	.8892
13.	−.8827	.3620	.3261	−.2909	1.1012
14.	.4667	.0780	.3170	−.3180	.4255
15.	−.0048	−.5407	.4750	.3947	.6738
16.	−.3475	.4927	.1914	−.2561	.4657
17.	.2249	.2966	−.2495	−.1099	.2129
18.	.6711	.3036	.1721	.3000	.6622
19.	.0792	.4571	.4304	−.1991	.4401
20.	−.2078	.3423	.3544	−.1533	.3094

C, *Y*, and *V*. The reader will recognize the similarity of this procedure to Eysenck's (1950) method of criterion analysis.

The final rotated factor loadings are shown in Table 4. With respect to the first hypothesis, Factor II′ is the one rotated so that its loadings on *F*% would be maximized. The hypothesis required that the loadings on *T*/*R* and *A*% be of the same sign as *F*% and that the loadings on *M*, *C*, *Y*, and *V* be of opposite signs. If we adopt a convention of considering only factor loadings of .35 or higher in order to deal with at least 10% of the variance of a score, the requirements of the hypothesis seem to be met. This factor resembles one interpreted as representing perceptual control by Wittenborn (cf. Factor B in 1950b). Such an interpretation is perhaps closer to the data than the one offered here. At the same time, it is interesting that the Rorschach can yield a factor predicted on the basis of a view of psychological efficiency as the central concept in mental health (Wishner, 1955).

Psychologically, this factor seems to reflect a continuum ranging from sensitivity to various nuances of the Rorschach cards (loadings of the same sign on *M*, *C*, *Y* and *V*, and *F*+) conforming to content most frequently reported (*F*+), at one end, to deliberate (*T*/*R*) and relatively exclusive attention to form (*F*%) with animal content (*A*%), at the other end. At a higher level of abstraction, one might interpret this as flexibility–rigidity. Taking into account previously demonstrated relationships with physiological variables, it does not seem too farfetched, even if somewhat speculative, to suppose that perceptual-flexibility-with-control may be the central basis for motor efficiency (Wishner: 1953, 1955). However speculative, it is still true that it was just this notion which led to the prediction apparently confirmed.

With respect to Hypothesis 2 concerning the factorial composition of *C*, *FC*, and *M*, rotations were made of Factors I′ and III, I′ and IV, and III and IV so as to maximize both *M* and *FC* alternately. None of these rotations produced a meaningful factor. It can be seen by an inspection of Table 4 that *M* and *FC* do not have the same factorial composition, as shown by the fact that they do not have significant loadings on the same factors, and indeed, they have loadings of opposite signs on two of the factors in the centroid analysis, and on one of the factors in the rotated analysis. At the same time, the

Table 4

Rotated Factor Loadings—Scoring Categories

Variable No.	I″	II′	III″	IV′	h^2
1.	−.6825	−.0207	.0012	.4181	.6410
2.	.7373	−.2606	.1603	.2544	.7019
3.	−.4149	.3055	−.1856	−.2278	.3518
4.	−.6611	.3021	.0795	.1113	.5470
5.	−.2228	−.4369	.6360	.1594	.6704
6.	.3594	−.5727	−.3293	.4224	.7440
7.	.4172	−.6527	−.4047	.5050	1.0189
8.	.2305	−.3429	−.2227	.1874	.2554
9.	.3983	−.6384	−.3859	.4644	.9308
10.	−.1503	−.6298	−.3859	−.1740	.5984
11.	−.1281	−.5726	.0788	−.3872	.5004
12.	−.2724	−.7884	−.2921	−.3280	.8887
13.	.0616	.9536	.1320	−.4121	1.1004
14.	.3190	−.3989	.0589	−.4012	.4253
15.	−.3794	−.2070	.6953	.0544	.6732
16.	.3522	.5123	−.0284	−.2793	.4653
17.	.2953	−.0909	−.3358	.0669	.2127
18.	.5663	−.4987	.1896	.2370	.6615
19.	.5366	.1058	.1754	−.3312	.4396
20.	.3073	.3250	.1768	−.2791	.3092

factorial composition of *FC* seems very similar to *C*, *C+CF*, *C+CF+FC*, and *Sum C/R*. Thus, this aspect of Wittenborn's results is not confirmed in this study.

With respect to Hypothesis 3, Factors I′ and III and Factors I′ and IV were rotated through *Dd%*, and this yielded Factor I″ which appears to be meaningful. The hypothesis required that there be high loadings on *Dd%*, *M%*, and *C*. This is seen, except that the loading on *M%* is not notably high. It is difficult, however, to exclude the possibility of artifact in that this may again represent some part of the productivity factor. Thus, for example, there is high loading on *R*, *V*, and *N*, as well as on the other locational variables.

The final rotation was an attempt to reproduce the productivity factor by rotating Factor IV against III′ through *R*. The productivity factor required that *R*, *W*, and *Z* have high loadings of the same sign, with *A* of opposite sign, based on the relationships with various subtests of the Wechsler-Bellevue found previously (Wishner, 1948). While this general pattern is seen in Factor IV′, the loadings do not appear high enough to warrant its interpretation as an intelligence factor; indeed, Factor I″ appears to fit the bill better. Thus, one might view both Factors I″ and IV′ as productivity factors, but having different psychological bases. Factor I″ may reflect productivity based on unusual fragmentations of the blots (note loadings on *D* and *Dd* contrasted with loadings of opposite sign on *W*, *P*, and *A*). The productivity reflected in Factor IV′ seems based on particular stimulability by color. Thus, we may have two independent types of productivity: the *Dd* type and the *C* type. Presumably, some optimal combination of these would reflect intelligence as measured by standard psychometrics, and this may account for the variable findings concerning intelligence on the Rorschach.

Factor III″ remains to be interpreted. It has significant loadings on *M* and *H*, and loadings of opposite sign on *C*, *Y*, and perhaps *N*. This would appear to indicate a tendency for individuals to predominate in one or the other, although there is no directly inverse correlation between *M* and *C* (see Table 1). In view of previously ambiguous findings in this sphere, it is unclear what psychological significance this factor may have.

Discussion

The results are not unlike those found previously, although a wholly different population from those in other studies was utilized. Almost all investigators since Wittenborn have found that the Rorschach yields four factors. Similarly, they have found a factor variously labeled as productivity, verbal intelligence, etc., while the other factors have proved ambiguous. The most thorough study by Eschenbach and Borgatta (1955) indicates the inadequacy of the Rorschach in the prediction of overt behavior.

In all these respects, the present study is confirmatory, or at least not contradictory. There is a failure to confirm Wittenborn's finding of similar factorial compositions for *FC* and *M*, which are different from *C* and *CF*; of course, no correlation with overt behavior was attempted.

On the positive side, there appears to be confirmation of a factor of psychological efficiency which would be related to mental health and illness as suggested in previous papers (Wishner: 1953, 1955). Clinically, after all, the Rorschach may well be conceived as having the estimate of mental illness as its major function, despite Rorschach's original denial that his test should have anything to do with psychiatric diagnosis. Thus, this confirmatory finding of a previous suggestion may be worthy of further investigation.

The h^2 column of Table 4 indicates that the greater part of the variance of *D%*, *FC%*, *F+%*, *A%*, *N/R*, *P%*, and *T/R* is unaccounted for by these factors. The first four of these have received extensive attention both in the Rorschach literature and in clinical work, but their significance, insofar as factor analyses can reveal it, seems in doubt. On the other hand, more than 70% of the variance of *W%*, *C* treated indifferently as between *CF* and *FC*, *Y+V*, and *F%* is accounted for. *F%* and the color and shading determinants may be conceived as two faces of the same coin. In general, the sign of the factor loadings of $F\%$ are the opposite of those of Y, V, and C. This is practically guaranteed by the scoring procedures, since high $C\%$, $Y\%$, and $V\%$ necessarily places a limit on $F\%$. It appears, therefore, that $F\%$ as a score reflecting perceptual control merits much greater attention than it generally receives, particularly in the light of the conception of psychological health here proposed. It is interesting that some 67% of the variance of $M\%$ is accounted for and that $F\%$ and $M\%$ tend to have similar factorial compositions. In view of past research on M (see particularly Singer, Wilensky, & McGraven, 1956; Singer & Spohn, 1954; Williams & Lawrence, 1954; Wittenborn, 1950b) this lends weight to the interpretation of $F\%$ offered here. At the same time, the factorial composition of $W\%$ seems independent of $F\%$ and may merit further attention as a possible indicator of a significant personality trait.

It must be recognized, of course, that failure to achieve simple structure implies a lack of uniqueness to the set of rotated factors chosen for interpretation. Certainly, other rotations testing other hypotheses would have been possible. At the same time, the final solution is not arbitrary, and rests on a body of previous research on the Rorschach and on aspects of personality theory.

First Response Times

The first response times recorded for Beck's sample were also factor-analyzed. In addition, first response times in the Rorschach test were correlated with response times to the Street Gestalt Completion Test (1931) and visual reaction times, studies suggested by the results of the factor analysis.

Previous research on response times in the Rorschach test has raised serious questions concerning the tenability of such concepts as "color shock," and "shading shock" (Dubrovner, Von Lackum, & Jost, 1950; Lazarus, 1949; Matarazzo & Mensh, 1952; Rabin & Sanderson, 1947; Siipola, 1950). There has been the positive suggestion that reaction time to the cards may be a function of the objective difficulty of each of the cards (Matarazzo & Mensh, 1952; Meer, 1955; Rabin & Sanderson, 1947). A search of the literature fails to reveal any previous factor analysis of first response times in the Rorschach test.

Table 5

Intercorrelations of First Response Times
to the Rorschach Cards
(Decimals omitted)

Card No.	No.: 1	2	3	4	5	6	7	8	9
2.	26								
3.	30	43							
4.	27	34	48						
5.	10	35	34	32					
6.	37	23	30	29	10				
7.	28	49	44	37	19	54			
8.	46	33	27	37	13	52	31		
9.	15	41	38	46	17	40	34	53	
10.	20	29	40	32	26	41	42	34	29

The object of this part of the study was to determine the number of factors necessary to account for the intercorrelations among first response times and to ascertain whether any such factors could be identified as difficulty or shock factors as had been suggested previously.

Procedures and Results

The raw data were the first response times to each of the ten Rorschach cards obtained from the sample described above, and the product-moment correlation coefficients among these are shown in Table 5. The results of the centroid analysis yielding two factors are shown in Table 6.[2] Although the centroid solution is arbitrary, a plot of the factors shown in Table 6 reveals immediately that while the structure might be simplified slightly by a small clockwise rotation of Factor I (10–15 degrees), the essential nature of the factors could not be altered by any sensible rotation. It was decided, therefore, to treat the centroid analysis as the final solution, without rotation.

Since there are relatively high loadings on all cards, it is apparent that Factor I reflects some general factor of individual differences in response time. Factor II is extremely difficult to interpret. It has probably significant

---

[2] For the major portion of this part of the statistical work, the author is indebted to Robert Downing.

Table 6

Centroid Factor Loadings—First Response Times

Card No.	I	II	h^2
1.	.48	−.22	.28
2.	.61	.22	.42
3.	.64	.29	.49
4.	.62	.09	.39
5.	.39	.35	.27
6.	.62	−.25	.45
7.	.66	.10	.45
8.	.64	−.46	.62
9.	.62	−.12	.40
10.	.56	.08	.32

loadings of opposite signs on Cards 5 and 8, but no immediately sensible interpretation is suggested, since Cards 5 and 8 do not have any uniquely polar relationship to each other which is phenomenally obvious or which has been suggested in the literature. It is interesting to note that the factors derived cannot account for much more than 50% of the variance of response times to these cards.

If Factor I is interpreted as a general factor, the problem arises whether it reflects general response speed of individual *Ss*, or whether it is peculiar to responses to inkblots, or perhaps even to the Rorschach inkblots in particular. To answer these questions the following experiment was performed.

The Relationship Between First Response Times on the Rorschach, Visual Reaction Time, and Response Times to Street Gestalt Completion Cards [3]

Thirty male college students were tested by means of the Rorschach test, the Street Gestalt test, and a Vernier chronoscope. Each *S* was given the Street Gestalt Completion Test first. Card A was used as a demonstration card to orient the *S* to his task. He was told to find the correct answers to the remaining cards as quickly as possible. He was given a ready signal and then a turn signal for each of the six cards, and his responses were timed by means of a stop watch.

The 10 Rorschach cards were then given to the *S* one at a time, with the same instructions as used for the Street Gestalt test, i.e., to give a response as quickly as possible, taking the whole inkblot into consideration. The *S* was not told that there were no correct answers in this test, since it was thought that this would make the situation somewhat less comparable to the Street Gestalt test.

Finally, 10 readings on a Vernier chronoscope were made of the *S*'s reaction time. *E* shielded his key from *S*'s view and told him to watch *E*'s pendulum. After a ready signal, the *S* was to release his pendulum as soon as he saw the other pendulum move. The need for speed was stressed.

Through the use of these tasks, it was hoped to compare the Rorschach with a conceptually oriented task, as well as with a visual-motor reaction time. Thus, if there were some general response speed factors involving conceptual activity, there should be significant correlations between the Rorschach and Street tests; if the general factor were confined to sensory-motor reaction times, there should be a correlation with visual reaction times; or, of course, the comparisons might yield intercorrelations among all three tasks, indicating a very general response time factor.

Results

Mean response times to the Street cards and to the Rorschach cards yielded a rho of +.23; the rho between reaction time and the Street Gestalt cards was −.03. Rhos between each of the 10 Rorschach cards and reaction time, and between each of the six Street Gestalt cards and reaction time, yielded no significant coefficients. It is clear, therefore, that no such general factor as cognition time or response time can be hypothesized to account for the general factor found in response times to the Rorschach test. It must be noted, of course, that the experimental procedure utilized with respect to response times on the Rorschach was unlike the clinical procedure; yet the instructions chosen, insofar as they were more homologous with the procedures in the visual reaction time and Street Gestalt experiments, should have favored higher intercorrelations among these tests. Factor I, therefore, appears to represent a factor more specifically related to inkblots and their effect on response times.

---

[3] The writer is indebted to William Rosenblith who aided in the collection of these data.

Summary

A matrix of intercorrelations of the scores on 157 Rorschach protocols originally collected from a stratified normal sample by Beck et al. (1950) was factor-analyzed. The hypotheses tested were: 1. That there is a factor of psychological efficiency; 2. That the factorial composition of *FC* is different from that of *C* and *CF;* 3. That there is a factor of verbal productivity; 4. That there is a factor of compulsivity. Support was adduced for Hypotheses 1 and 4; Hypothesis 3 also seems tenable, but with less confidence; Hypothesis 2 is not confirmed. Perceptual control, as reflected in *F%* and *M%*, seemed to be the most important variable involved in the conception of psychological health on the Rorschach test.

A factor analysis of the intercorrelations among the first response times to the 10 cards of the Rorschach test yielded two factors, one of which seemed interpretable as a general timing factor. In order to determine its degree of generality, 30 subjects were given the Rorschach and Street Gestalt tests, and a visual reaction time task. No significant intercorrelations were found. It was concluded, therefore, that the general factor seen in the response times of the Rorschach test cannot be conceptualized as a generalized cognition time factor, but is probably related more specifically to inkblots and their specific effects on response times.

Received August 11, 1958.

REFERENCES

Beck, S. J., Rabin, A. I., Thiesen, W. C., Molish, H., & Thetford, W. N. The normal personality as projected in the Rorschach test. *J. Psychol.*, 1950, **30**, 241–298.

Borgatta, E. F., & Esehenbach, A. E. Factor analysis of the Rorschach variables and behavioral observation. *Psychol. Rep.*, 1955, **1**, 129–136.

Consalvi, C., & Canter, A. Rorschach scores as a function of four factors. *J. consult. Psychol.*, 1957, **21**, 47–51.

Cronbach, L. J. Statistical methods applied to Rorschach scores. *Psychol. Bull.*, 1949, **46**, 393–429.

Dubrovner, R. J., Von Lackum, W. J., & Jost, H. A. A study of the effect of color on productivity and reaction time in the Rorschach test. *J. clin. Psychol.*, 1950, **6**, 331–336.

Eschenbach, A. E., & Borgatta, E. F. Testing behavior hypotheses with the Rorschach: An exploration in validation. *J. consult. Psychol.*, 1955, **19**, 267–273.

Eysenck, H. J. Criterion analysis—an application of the hypothetico-deductive method to factor analysis. *Psychol. Rev.*, 1950, **57**, 38–53.

Fiske, D. W., & Baughman, E. E. Relationships between Rorschach scoring categories and the total number of responses. *J. abnorm. soc. Psychol.*, 1953, **48**, 25–32.

Holtzman, W. H., & Sells, S. B. Prediction of flying success by clinical analysis of test protocols. *J. abnorm. soc. Psychol.*, 1954, **49**, 485–490.

Lazarus, R. S. The influence of color on the protocol of the Rorschach test. *J. abnorm. soc. Psychol.*, 1949, **44**, 506–516.

Matarazzo, J. D., & Mensh, I. N. Reaction time characteristics of the Rorschach test. *J. consult. Psychol.*, 1952, **16**, 132–139.

Meer, B. The relative difficulty of the Rorschach cards. *J. proj. Tech.*, 1955, **19**, 43–53.

Rabin, A. I., & Sanderson, M. H. An experimental inquiry into some Rorschach procedures. *J. clin. Psychol.*, 1947, **3**, 216–225.

Sen, A. A statistical study of the Rorschach test. *Brit. J. Psychol.*, 1950, **5**, 21–59.

Siipola, Elsa M. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358–382.

Singer, J. L., & Spohn, H. E. Some behavioral correlates of Rorschach's experience-type. *J. consult. Psychol.*, 1954, **18**, 1–9.

Singer, J. L., Wilensky, H., & McGraven, V. G. Delaying capacity, fantasy, and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.*, 1956, **20**, 375–383.

Street, R. F. A Gestalt completion test: A study of a cross-section of intellect. *Teach. Coll. Contrib. Educ.*, 1931, No. 481.

Williams, H. L., & Lawrence, J. F. Further investigation of Rorschach determinants subjected to factor analysis. *J. consult. Psychol.*, 1953, **17**, 261–264.

Williams, H. L., & Lawrence, J. F. Comparison of the Rorschach and MMPI by means of factor analysis. *J. consult. Psychol.*, 1954, **18**, 193–197.

Wishner, J. Rorschach intellectual indicators in neurotics. *Amer. J. Orthopsychiat.*, 1948, **18**, 265–279.

Wishner, J. Neurosis and tension: An exploratory study of the relationship of physiological and Rorschach measures. *J. abnorm. soc. Psychol.*, 1953, **48**, 253–260.

Wishner, J. The concept of efficiency in psychological health and in psychopathology. *Psychol. Rev.*, 1955, **62**, 69–80.

Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1950, **14**, 261–276. (a)

Wittenborn, J. R. Level of mental health as a factor in the implications of Rorschach scores. *J. consult. Psychol.*, 1950, **14**, 469–472. (b)

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

SOME MANIFEST CHARACTERISTICS OF RECALLERS AND NONRECALLERS OF DREAMS

ROSALEA A. SCHONBAR
Teachers College, Columbia University

Dreaming, most psychologists agree, is a universal experience. Research with electrically recorded eye movements during sleep led Dement and Kleitman (1957) to conclude that "periods of . . . dreaming . . . are an intrinsic part of normal sleep" (p. 345), and that all people dream several times a night.

Although it is likely that everyone dreams, many people report that they do not. The fact, then, that many persons have little or no recall of dreams raises interesting questions. What are the distinguishing characteristics of persons who do, or do not, report that they dream?

The present research was undertaken to examine some correlates of dream recall. Some hypotheses were drawn from Ramsey's (1953) review of the literature on dreaming—that frequent recallers of dreams are younger, more intelligent, and more often women than men. Other hypotheses were derived from psychoanalytic theories (Fromm, 1951; Hadfield, 1954) which hold that the dream makes possible an internal communication prohibited during consciousness because of the anxiety it would evoke. Two alternative propositions were put to test: that manifest anxiety is positively related to the recalling of dreams, or that nonrecallers and frequent recallers are both more anxious than a less extreme group.

The informal observation that many psychoanalytic patients become progressively better able to remember dreams leads to the following question: to what extent is this because remembering is "the thing to do" under the circumstances, and to what extent because it implements their motivated search for self awareness? Manifest needs were selected from those defined by Edwards (1954), and the following predictions made. If conformity is the determining factor, then people who recall dreams when it is asked for or expected of them might be expected to have high needs for achievement, deference, and authority, and a low autonomy need; if, on the other hand, a search for self awareness motivates dream recall, recallers might be expected to have high needs for endurance, intraception, and succorance.

Procedure

The *S*s were 42 teachers and school guidance counselors who were students in a six-week graduate summer course in parent counseling. There were 15 white women, 13 Negro women, and 14 white men. Information concerning age was collected in terms of five-year intervals; taking midpoints, the age range was from 22 to 52 years, with the mean at 33.8 years. Men and women did not differ significantly in age, but only 32% of the women were or had been married, in contrast to 62% of the men.

On the first day of class a personal data form was distributed. These sheets contained code numbers, which were thereafter the only identifications on material submitted for the study. Each *S* was given a printed booklet containing a detailed instruction sheet and 28 perforated pages.

The instructions stated that one page was to be submitted each day for 28 days, containing a record of all dreams or experiences of dreaming while asleep during the preceding 24 hours. *S*s were told to write the dreams as completely as possible, including incongruities, vague impressions, and the like; they were told how to report associations or clarifications so as not to contaminate the dream report itself. Feelings during and after the dream were to be recorded; *S*s were also instructed to indicate whether the dream occurred while falling asleep, at an indeterminate time, or just before waking, or if it woke them during the night. Dreams were to be written immediately upon waking in the morning, or on waking from a nap.[1]

Each report sheet contained a short reminder of some of the more important instructions. The rest of the page was blank so that the dream could be recorded. At the bottom of each sheet *S*s could check, if appropriate, "No dreams in this interval," "Aware of having had dream but can't remember content," or "Above (recorded dream) is all I remember, but I know there was more." Each *S* submitted a page every day. In tabulation, an *S* was credited with having remembered a dream if he could remember any fragment of content. Although some *S*s reported more than one dream during a single report period, frequency of recall was defined for this study as the number of report periods in which one or more dreams were recalled.

During the second week of the experiment the Edwards Personal Preference Schedule (1954) was given to the *S*s to complete at home. Early in the fourth week the IPAT Anxiety Scale (Cattell, 1957) and a 20-word vocabulary test (Thorndike, 1942) were administered in class. Briefly, the Edwards is an inventory of paired statements, matched for social acceptability, in which the forced-choice scores correspond to a need system based on the work of H. A. Murray. The IPAT Anxiety Scale is a 40-item questionnaire based on factor analysis of a number of measures, and yielding split-half reliabilities of .84 and .91 on different populations; validity studies have also yielded positive results. The multiple-choice vocabulary test was derived from the IER Intelligence Scale CAVD; it correlates .50 with a general intelligence factor among a sample of adult males (Thorndike, Norris, & Morrill, 1952) and .62 with the 1916 Binet administered to hospital patients (Miner, 1952).

Results

In all, 1176 reports were collected, of which 215 contained dreams. The range of individual recall frequency was from 0–19 days; the mean frequency was 5.1, the median 4.0, indicating that half the *S*s recalled dreams only 14% of the time or less. Fifteen *S*s recalled dreams only once or not at all; these constitute the Nonrecallers. Thirteen of the *S*s recalled dreams 25% of the time or more; these constitute the Recallers. The Recallers consisted of four men and nine women, of whom four were Negroes; the Nonrecallers contained five men and ten women, of whom five were Negroes.

Sex and race. The mean frequency of dream recall for men was 4.57, and for women 5.39; while in the predicted direction, the difference (measured by t test) was not significant. The mean for Negro women was 5.0, for white women 5.73, a nonsignificant difference. There was no sex difference in the constituency of the two groups.

The means and standard deviations of the Recallers and Nonrecallers were computed on all variables; the differences between the group means were measured by t tests. Table 1. presents these data. When a difference was significant, the point biserial correlation is also presented to indicate the degree of relationship between the variable and dream recall. All comparisons are for the two extreme groups only.

Age and manifest needs. As Table 1 reveals, age is not related to frequency of dream recall. Of the manifest needs, the only significant difference was for succorance; five of the other six hypothesized differences were in the predicted direction, but were insignificant. It seems most economical to conclude that the difference in succorance was a chance finding, and that there is no relationship between these needs as tested and recalling or not recalling dreams.

Intelligence and manifest anxiety. Table 1 reveals that Recallers were significantly higher than Nonrecallers both in intelligence and in anxiety. The mean anxiety score of the middle group ($N = 14$) in dream recall was 27.07,

---

[1] The detailed instruction sheet has been deposited with the American Documentation Institute. Order Document No. 6015, remitting $1.25 for microfilm, or $1.25 for photocopies.

Table 1

Mean Differences Between Recallers and Nonrecallers in Age, Vocabulary, Manifest Needs, Manifest Anxiety, and Contentless Recall

Variable	Recallers ($N = 13$)		Nonrecallers ($N = 15$)				
	M	SD	M	SD	t	p[a]	r_{pbs}
Age, years	36.84	13.13	34.33	7.94	.72	n.s.	—
Vocabulary	15.15	3.46	12.53	3.91	1.79	.05	.35
Manifest anxiety							
Total	34.92	9.05	25.40	9.86	2.81	<.01	.48
Overt	18.38	6.67	11.46	5.66	2.86	.01	.50
Covert	16.54	3.08	13.93	5.40	1.46	n.s.	—
Self sentiment[b]	6.38	2.68	5.00	2.76	1.19	n.s.	—
Ego strength[b]	5.00	1.88	2.06	2.08	3.77	<.01	.59
Guilt proneness	11.38	3.16	9.13	3.44	1.79	n.s.	—
Ergic tension	8.08	3.39	6.40	3.14	1.28	n.s.	—
Manifest needs[c]							
Achievement	14.54	3.73	13.86	2.98	.51	n.s.	—
Deference	15.62	3.91	13.64	3.04	1.43	n.s.	—
Exhibitionism	10.00	4.19	12.72	2.79	—[d]	—	—
Autonomy	10.77	2.83	11.36	3.02	.51	n.s.	—
Intraception	19.23	4.14	18.86	3.30	.25	n.s.	—
Succorance	14.53	4.07	11.07	5.18	1.87	.05	.35
Endurance	14.62	5.69	14.57	5.10	.02	n.s.	—
Contentless recall	8.85	3.19	1.87	2.75	5.79	.001	.75

[a] One-tailed tests were applied to the predicted differences, two-tailed tests to the subcategories of manifest anxiety and to contentless recall.

[b] High scores indicate low self-sentiment and ego strength.

[c] For these variables, $N = 14$ for Nonrecallers, as one S did not complete the schedule.

[d] Hypothesis not confirmed; hence, difference not tested.

falling between rather than below the means of the Recallers and the Nonrecallers; any test, then, of the hypothesis that this group would have lower mean anxiety than the extreme groups is unnecessary. However, in order to determine whether the relationship between the variables was significant over the whole distribution, the correlation ratio was computed and equaled 4.53, significant at better than the .05 level. It may be concluded that there is a positive relationship between manifest anxiety and dream recall.

Discussion

The findings of this study are, of course, limited because of the relatively small number of Ss in each of the groups. Nevertheless, some of the findings are comparable to those of some earlier studies, and, where there are discrepancies, other factors may be responsible, such as methodological differences and the fact that the present sample is somewhat older than many of the other samples, as well as more homogeneous in occupation and more heterogeneous in race. That the intelligence and educational levels of the group were relatively high—the mean vocabulary score of the present sample was 13.84, equivalent to approximately the 90th percentiles of other groups tested (Hagen & Thorndike, 1955) and to the 73rd percentile of the ACE (Miner, 1957)—limits the generality of the findings, but does not significantly differentiate this sample from previous samples in related studies.

There was a striking variation from day to day in the number of people recalling dreams; the range was from 2 to 17 a day. A study of the Recallers' daily variations revealed that there was no significant trend for different days of the week. Coefficients of correlation were also computed between the number of nocturnal dream reports from the Recallers each day and the average temperature, humidity, and humiture for the corresponding period (10:00 p.m. to 7:00 a.m.);[2] the correlations were .16, .01, and .06, respectively. Previous studies have not accounted for either individual or group variability in recall over time. The present study has eliminated certain factors, but leaves the observation unexplained.

The greatest deviation from earlier findings is that neither sex (McElroy, 1952; Middleton, 1942) nor age (Kleitman, 1939) was associated with ability to recall dreams. In view of this discrepancy, the present sample was compared with those of Middleton (1933; 1942) to determine whether recall productivity itself differed to any great extent. In spite of the fact that Middleton's findings were based on responses to a questionnaire requiring no documentation, the percentages of his *S*s reporting experiences of dreaming and frequent or very frequent dreaming are in fairly close agreement with percentages of the present sample reporting at least one dream, or being categorized as Recallers, respectively. Compared with reports by Kleitman (1939), the present sample was not atypical in the percentage stating that they never dream. The finding that Recallers score higher in verbal intelligence confirms earlier findings (Ramsey, 1953), although the relationship was not a major one; the fact that dream reports were written, thus posing greater demands both motivationally and operationally upon the less verbally adept, suggests that, notwithstanding the earlier findings, the relationship may be partly artifactual.

The results concerning the more dynamic factors of manifest needs and manifest anxiety raise interesting theoretical questions, particularly with respect to the unknown relationship between manifest and covert manifestations of the same characteristic. The need variables, for example, were initially thought of in two categories. In the first, thought of as situational, it was predicted that individuals who had needs to do well (achievement), to "show off" (exhibitionism), and to please authority (deference and low autonomy) might be motivated to recall dreams because it was a classroom instructor who was asking them to do so; in a more pervasive motivational category, it was predicted that people with stronger needs to understand themselves (intraception), to seek help (succorance), and to stick with problems (endurance) would be found among the frequent Recallers, in accordance with the theory that remembering dreams offers a unique opportunity for insight (Fromm, 1951), creativity (Murphy, 1947), and problem solution (Hadfield, 1954). The negative findings leave open the question raised initially concerning the motivation for increased recall of dreams during psychoanalytic therapy. It may be that the instrument used is not a valid one, or that needs with the names and characteristics studied here are related to dream recall but not at the manifest level.

Similarly, the finding that manifest anxiety plays a substantial role in dream recall can mean either that anxious people are more urgently pressed to resolve the conflicts that dreams theoretically illustrate, that they have more such conflicts, or that persons who have erected fewer barriers between themselves and awareness of their anxiety are also more fully in touch with the rest of their internal experience and, hence, recall dreams more frequently. The latter view receives some support from the fact (Table 1) that the groups differed in overt but not covert anxiety, and that they did not differ in unresolved tensions (ergic tensions), but that the Nonrecallers were more able to "control and express" them realistically (Cattell, 1957, p. 5).

The high correlation between remembering dreams and the contentless recall of dreaming also supports the view that low dream recall, and possibly low manifest anxiety, is related to repression. While it is obvious that, in order to recall dreams, one must experience dreaming, it does not necessarily follow that those who do not recall dreams would also not remember dreaming, especially since the latter is an "easier" task, and the two categories were mutually exclusive. It seems likely that the almost absolute lack of recall of dreaming itself is due to factors associated with repression or control, and that it is therefore symptomatic of a more general lack of awareness of ongoing internal processes.

---

[2] These data were obtained from the United States Weather Bureau. Humiture is a comfort-discomfort measure obtained by averaging temperature and humidity for a given time period.

Summary

Forty-two graduate students in education turned in reports on recalled dreams every day for four weeks. They also completed the Edwards Personal Preference Schedule, the IPAT Anxiety Scale, and a short multiple-choice vocabulary test. On the basis of the frequency with which they recalled dreams, subjects were divided into a group of Recallers and a group of Nonrecallers. By means of t tests applied to differences between the means of the subgroups on the variables tested, the following conclusions were reached:

1. Men and women do not differ in frequency of dream recall. Small sample size may be a factor here, since the nonsignificant difference was in the predicted direction.

2. Dream Recallers are not younger than Nonrecallers within the age range covered in this study.

3. Dream Recallers are more intelligent than Nonrecallers.

4. There is a positive relationship between manifest anxiety and frequency of dream recall. The findings also suggested that the difference between Recallers and Nonrecallers was in overt rather than covert anxiety, and that, although they do not differ in unresolved tensions, Recallers have less ego strength than Nonrecallers.

5. There is no relationship between the frequency of recalling dreams and manifest needs for achievement, deference, exhibitionism, intraception, succorance, endurance, or autonomy.

6. There is no relationship between diurnal variations in dream recall and variations in temperature, humidity, or humiture.

7. Contentless recall of dreaming is positively related to recall of dreams.

The study was based upon the assumption that dreaming is a universal process. On the basis of the present findings, variations in ability to recall dreams or dreaming were discussed in terms of a repressive factor operating most successfully in total Nonrecallers. Attention was called to the limitations of the study.

Received August 15, 1958.

REFERENCES

Cattell, R. *Handbook for the IPAT Anxiety Scale (self analysis form)*. Champaign, Ill.: Instit. Personality & Ability Testing, 1957.

Dement, W., & Kleitman, N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *J. exp. Psychol.*, 1957, **53**, 339–346.

Edwards, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

Fromm, E. *The forgotten language.* New York: Rhinehart, 1951.

Hadfield, J. A. *Dreams and nightmares.* London: Penguin, 1954.

Hagen, Elizabeth, & Thorndike, R. L. Normative test data for adult males obtained by house-to-house testing. *J. educ. Psychol.*, 1955, **46**, 207–216.

Kleitman, N. *Sleep and wakefulness.* Chicago: Univer. Chicago Press, 1939.

McElroy, W. A. The frequency of dreams. *Quart. Bull. Brit. psychol. Soc.*, 1952, **3**, 91–94.

Middleton, W. C. Nocturnal dreams. *Sci. Mon.*, 1933, **37**, 460–464.

Middleton, W. C. The frequency with which a group of unselected college students experience color dreaming and color hearing. *J. genet. Psychol.*, 1942, **22**, 221–229.

Miner, J. B. *Intelligence in the United States.* New York: Springer, 1957.

Murphy, G. *Personality.* New York: Harper, 1947.

Ramsey, G. V. Studies of dreaming. *Psychol. Bull.*, 1953, **50**, 423–455.

Thorndike, R. L. Two screening tests of verbal intelligence. *J. appl. Psychol.*, 1942, **26**, 128–135.

Thorndike, R. L., Norris, R., & Morrill, C. S. General aptitude test battery scores in a regional sample. *USAF Hum. Resour. Res. Cent., Res. Note,* 1952, No. 52-16.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

INTERRELATIONSHIPS AMONG MMPI MEASURES OF DISSIMULATION UNDER STANDARD AND SOCIAL DESIRABILITY INSTRUCTIONS[1]

JERRY S. WIGGINS
Stanford University

Current studies of what Jackson and Messick (1958) have recently termed "stylistic" aspects of responses to personality inventories have devoted considerable attention to the "social desirability" component (Corah, Feldman, Cohen, Gruen, Meadow, & Ringwall, 1958; Edwards: 1954, 1957; Fordyce, 1956; Hanley: 1956, 1957; Hillmer, 1958; Rosen, 1956; Voas, 1956; Wiggins & Rumrill, 1959). Social desirability response style may be defined as a general tendency to endorse personality inventory items that are judged to be socially acceptable by people in general (Edwards, 1957). Measurement of this response style has been indirect, at best. The standard method of demonstrating its presence has been to correlate item endorsement frequency with independently rated item social desirability in groups of subjects taking a personality inventory (Edwards, 1957; Hanley, 1956). Such studies have demonstrated that about 76% of the variance involved in group endorsement may be accounted for by item social desirability values alone. This approach provides little information about the extent to which *individuals* are answering in terms of item social desirability, and the reported relationship appears to be so omnipresent among self-report devices as to be of little practical significance (Corah et al., 1958; Hillmer, 1958; Jackson & Messick, 1958).

A more direct approach to the measurement of this response style is represented by the attempts to develop social desirability dissimulation "scales" which, in principle, serve to identify individual dissemblers (Edwards, 1957; Hanley, 1957; Wiggins & Rumrill, 1959). Although not lacking in face validity or theoretical relevance, these scales lack evidence regarding their empirical (predictive) validity. Thus, the social desirability scale of Edwards (1957) (hereafter called *En*[2]) was constructed from rational sortings of 10 judges. Evidence of its validity is its high correlations with the MMPI clinical scales and the *K* scale (Fordyce, 1956) and its low correlations with the Edwards (1954) EPPS scales which are presumably free of social desirability. The scale of Hanley (*Ex*) is more elaborately derived, although its validity again appears to rest on correlations with the MMPI scales, the *K* scale and the *En* scale (Hanley, 1957).

The validity of these scales becomes an important consideration when the argument is reversed and the scales are used to show the susceptibility of the MMPI to social desirability response style (Edwards, 1957; Fordyce, 1956; Hanley, 1957) or the relative freedom of the EPPS from such tendencies (Edwards: 1954, 1957). The possibility exists that these scales (including *K*) are not adequate empirical measures of social desirability and cannot be considered as useful operational definitions of this tendency. In addition, there are reasons to believe that other stylistic components, such as response acquiescence, may be involved in these scales so that interpretations of their measurement properties must be highly qualified (Jackson

---

[1] This investigation was supported in part by a research grant, M-2215, from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

[2] This scale is usually called SD but is here referred to as *En* to indicate the 39-item version and to avoid confusion with other scales in the present study.

& Messick, 1958; Wiggins & Rumrill, 1959). These possibilities may be investigated by recourse to the method of contrasting groups which estimates the actual efficiency of these scales in identifying social desirability dissemblers.

The method of contrasting groups has previously been used to estimate the efficiency of existing MMPI measures of "defensiveness" and "plus-getting" (Cofer, Chance, & Judson, 1949; Gough: 1950, 1954; Grayson & Olinger, 1957; Hunt, 1948; McKinley, Hathaway, & Meehl, 1948; Meehl & Hathaway, 1946; Rosen, 1956; Schmidt, 1948) to estimate the fakability of certain scales (Benton, 1945) and to develop new measures of "dissimulation" (Gough, 1954) and "defensiveness" (Cofer et al., 1949). It would therefore seem to be the method of choice for evaluating current measures of "social desirability" response tendencies. The traditional procedure in these "role-playing" studies has been to contrast the protocols of small groups of subjects who have taken the MMPI under standard and one or more "fake" instructions. That this procedure introduces elements not present in the population to which results are generalized is suggested by the studies of Grayson and Olinger (1957) and Voas (1956). Grayson (1957) reports patients' comments such as: "Well, I just put down the opposite of what I did yesterday" (p. 75). Voas (1956) presents clear evidence that contiguity in time between self and "socially acceptable" answering tends to increase honesty of self-report—which would effect a counterbalanced design. The present study was designed with the recognition that the population about which we would generalize contains subgroups of high dissimulators and relatively honest people—rather than people taking the MMPI under several conditions. It therefore employs the method of contrasting two large groups randomly selected from a homogeneous population.

An appropriate criterion group for social desirability faking is difficult to obtain for, as Edwards (1957) has pointed out, the extent of social desirability faking in a "normal" group is unknown. This shortcoming is not peculiar to social desirability response style, however, and it seems reasonable to assume that a relatively uniform group of high social desirability responders may be created by special instructions to the subjects. Such a group would differ in the amount of this tendency and could, in principle, serve as a criterion group for distinguishing an unusual amount of social desirability response style from a small amount. Scales which purport to measure social desirability response style should presumably achieve a certain amount of success in differentiating such a group from a "normal" group.

The present study is concerned with the empirical success of several MMPI measures of "social desirability," "defensiveness," "dissimulation," and "response bias" in distinguishing groups of social desirability role-players from comparable control groups. Attention is also focused on the interrelationships of these measures under both dissimulation and standard conditions.

Method

The full scale MMPI was administered to 440 undergraduate men and women under one of two sets of instructions. The 190 *S*s in the control group received the standard set of instructions that is printed on the front of the booklet. The 250 *S*s in the experimental group received a modified set of instructions that was stapled over the standard instructions on the front of the booklet. The modified instructions read, in part:

> Read each statement and decide whether *People in General* would consider a true or a false answer to be more desirable. You are *not* asked whether the statement is true or false as applied to you. Rather you are asked to decide which answer you think *People in General* would consider to be more desirable.

These instructions were supplemented by an explanation that the judgment to be made was of the general values of American culture rather than of any particular subgroup and that the hypothetical respondent was to be considered as of the same sex as the judge.

The protocols were scored for the standard MMPI clinical scales, including Welsh's factor scales. In addition, 11 special scales were scored which are presumably measures of dissimulation, social desirability, or response bias. These scales are briefly described below.

L, F, K: These are the well-known validity scales of the MMPI (Meehl & Hathaway, 1946). *L* is a 15-item rational measure of *Ss'* tendencies to give an unrealistically favorable picture of themselves. *F* is a 64-item rational measure of malingering which consists of items scored in a direction that is rarely chosen by normals. *K* is a 30-item empirical measure of "defensiveness" or "minus-getting." Twenty-two items were selected which differentiated patients with high *L* scores and normal profiles from a comparable group of patients with abnormal profiles. To these items were added eight items which remained unchanged under role-playing instructions in normal groups and which also differentiated severely disturbed patients from normals.

En: Refers to the 39-item rational social desirability scale developed by Edwards (1957) and Fordyce (1956) and usually designated SD. Ten judges were instructed to answer 149 items (from the *L, F, K,* and Taylor Anxiety scales) in such a way as to give the most socially desirable picture of themselves (Edwards, 1958). There was unanimous agreement on 79 items which were subsequently reduced to 39 items by item analysis. Items are keyed in the desirable direction.

Ex: Refers to the 26-item rational social desirability scale developed by Hanley (1957). A pool of 53 items (which were endorsed by 36 to 64% of Hathaway's college students) was given to 92 judges to rate on a 9-point scale of social desirability. A statistical criterion of high- and low-rated items was set, and the 26 items that survived were keyed in the desirable direction.

Sd-A, Sd-R: Refer to the 39- and 40-item rational social desirability scales developed from Welsh's Factor Scales *A* and *R* (Wiggins & Rumrill, 1959). The combined pool of 79 items was rated for both "true" and "false" responses on a 7-point scale by a total of 181 judges. *Sd-A* is a pool of low-rated items, and *Sd-R* is a pool of neutral or moderate items. Both scales are keyed in the desirable direction.

Cof: Refers to the 34-item empirical "lie" scale developed by Cofer and associates (1949). This scale consists of those reliable items which were unchanged under fake-bad instructions but were changed under fake-good instructions in a design utilizing a total of 81 *Ss*. Two judges agreed that this item pool represents statements of low social desirability value that are frequently endorsed. Items are keyed in a desirable direction. Cross-validations have not been reported.

Ds: Refers to the 74-item empirical "dissimulation" scale developed by Gough (1954) in his study of misconceptions about neuroticism. This scale consists of those items which differentiate genuine neurotics and normals from *Ss* instructed to fake neuroticism. A group of 111 dissemblers was contrasted with 176 patients. Items are keyed in the direction of neurotic dissimulation. Extensive cross-validational data are reported.

B: Refers to the 63-item rational "response bias" scale developed by Fricke (1957). This scale consists of those non-*K* items endorsed by 40 to 60% of Hathaway's combined college and normal groups. These items are arbitrarily keyed "true" and provide a measure of acquiescent or response-bias tendencies.

Sd: Refers to a 40-item empirical social desirability scale developed in the course of the present study. This scale consists of those items which discriminate *Ss* instructed to answer in a socially desirable direction from a comparable control group. A group of 178 dissemblers was contrasted with 140 controls. Items are keyed in a socially desirable direction. Cross-validational data is available.[3]

Results and Discussion

Clinical Scales

Comparisons of mean raw scores on the clinical scales between the dissimulation and control groups were made separately for men and women by means of *t* tests. In contrasting the mean scale scores of the 144 experimental men with the 105 control men, significant reductions at the .001 level were found on *Si*, *Hy*, and *R* (in order of decreasing magnitude). At the .01 level, a decrease occurred on *Pt* and an increase on *Ma*. Welsh's Scales *A* and *D* were found to have decreased significantly at the .05 level. To the college male, social desirability seems to involve social extraversion, high activity, and an absence of anxiety and somatic complaints. (The last inference comes from the fact that *Hy-subtle* scores tended to be almost identical under the two conditions, while reliable differences occurred on *Hy*.)

Larger and more numerous differences were found in contrasting the 106 experimental women with the 85 control women. Significant reductions in mean score were found on *Si, Mf, Pt, A, R, Hy,* and *Hs*—all at less than .001. At the .01 level, decreases occurred on *D, Sc,* and *Pd*. College women seem to ascribe more stringent and widespread criteria of social desirability to our culture (as our culture does to college women). Again we find social extraversion and an absence of anxiety and somatic complaints as socially desirable—this time without the associated high activity level. More pronounced among women than among men are the decreases in deviant thinking and deviant social attitudes found in *Sc* and *Pd*. The latter two were also tendencies in men,

---

[3] Description of scale development is reserved for a later publication.

Table 1

Mean Raw Scores on Dissimulation Measures Under Standard and Social Desirability Instructions

Group		*L*	*F*	*K*	*En*	*Ex*	*Sd–A*	*Sd–R*	*Cof*	*Ds*	*B*	*Sd*
Experimental Men ($N = 144$)	Mean	6.48	5.21	16.85	32.51	16.52	30.34	29.73	20.25	10.83	28.01	24.08
	SD	4.07	4.52	5.87	5.68	5.13	7.28	4.69	7.29	8.07	6.28	7.53
Control Men ($N = 105$)	Mean	2.80	5.24	16.33	30.99	13.25	27.81	26.91	12.93	10.52	26.46	12.06
	SD	1.76	3.60	4.55	4.95	3.32	7.18	3.75	3.59	6.19	6.12	3.59
	t	8.51	.23	.76	2.20	5.70	2.72	5.08	9.47	.33	1.94	17.89
	P	.001			.05	.001	.01	.001	.001			.001
Experimental Women ($N = 106$)	Mean	7.41	4.29	17.74	33.33	17.90	31.49	30.79	21.43	9.40	28.25	27.13
	SD	3.82	3.80	5.54	5.56	4.17	6.76	3.30	6.21	7.40	6.67	6.69
Control Women ($N = 85$)	Mean	3.07	3.67	15.59	29.40	12.91	25.65	27.41	11.19	10.45	27.31	12.41
	SD	2.09	2.52	4.23	4.82	2.98	7.03	2.72	3.72	5.31	5.81	3.97
	t	9.41	1.29	2.97	5.19	9.28	5.82	7.59	13.38	1.10	1.02	17.90
	P	.001		.01	.001	.001	.001	.001	.001			.001

although not statistically reliable ones. A striking difference between experimental and control women which did not occur between men was found on the *Mf* scale. By implication, college women tend to see themselves as more masculine than our culture would have them.

These clinical scale differences are highly similar to those reported by Rosen (1956) in a situation in which the same *S*s responded in terms of self-rating and social desirability. Earlier reports of normal *S*s faking good on the MMPI (Cofer et al., 1949; Hunt, Carp, Winder, & Kantor, 1948; McKinley et al., 1948; Meehl & Hathaway, 1946) tended to minimize scale differences. They emphasized that the fake-good *profile* is slightly lower, but of such similar shape that it cannot be discriminated from other normal profiles. The present findings (as well as those of Rosen) in no way contradict this earlier statement.

Validity Scales

Table 1 presents the mean raw scores on the dissimulation measures under standard and role-playing instructions for men and women treated separately. The reliable differences in *L* scale scores for both men and women are in accord with those reported by Cofer et al. (1949) and by Rosen (1956). Mean *F* scores under the two conditions were virtually identical for men and highly similar for women, which further confirms the observation that *F* is of no value in detecting positive malingering (Cofer et al., 1949; Gough, 1950; Hunt, 1948; Meehl & Hathaway, 1946). Although a reliable difference in mean *K* raw score occurred between the groups of women, *K* scores were almost identical in the two groups of men. The additive combination of *L* and *K* suggested by Cofer would contribute nothing further to the detection of male dissemblers in the present study. Cofer reports differences in *K* for a mixed group, and Rosen found differences for the sexes treated separately. However, an earlier study reported in McKinley et al. (1948) found no differences in *K* for either of the sexes treated separately. McKinley suggests that no differences were found on *K* because of some factor in the experimental situation that resulted in exceptionally high *K*s under standard conditions. Apparently, this factor was not operative in the present design, since the raw *K* scores under control conditions for Stanford men and women (Table 1) do not significantly depart from the norms given by McKinley for college men and women ($\bar{X} = 16.10$, $\sigma = 5.15$; $\bar{X} = 15.58$, $\sigma = 4.20$). The actual screening efficiency of *L* and *K* in identifying college dissemblers is discussed in the next section.

Special Scales

The remaining columns of Table 1 show the special scale differences under standard and role-playing instructions for both sexes. For the men, *Sd, Cof, Ex,* and *Sd-R* show a reliable difference at the .001 level. *Sd-A* was different at the .01 level and *En* at the .05 level. The response bias scale (*B*) approached, but did not attain, significance in this group. For the women, *Sd, Cof, Ex, Sd-R, Sd-A,* and *En* were all reliably different at the .001 level.

The actual screening efficiency of these measures was assessed by combining the samples into a group of 250 dissimulators and a group of 190 controls and calculating the proportion of hits and misses at various cutting scores for each scale. Space limitations prohibit the presentation of the complete distribution for each of the scales. However, Table 2 shows the degree of successful classification involved at the cutting score which minimizes *both* false positives and false negatives.

A certain amount of overlap between the distributions of dissemblers and controls is to be expected on the grounds that an as yet unknown proportion of dissemblers will be found in any control group. However, it is clear from Table 2 that a group of known (instructed) dissemblers may be differentiated from such a control group with better than chance success. The advantage of empirical scales in this enterprise is also apparent. The success of *Sd* in identifying dissemblers is, no doubt, somewhat spurious in that the scale was developed on part (72%) of the present sample. To test the extent of spurious inflation, *Sd* was scored in *independent* samples of 50 control college men and 72 college men instructed to answer in terms of social desirability. At a cutting score of 21 or above, the scale successfully identified 68% of the dissemblers and 100% of the controls (phi = .683). The success of *Sd* in the present study is therefore more than suggestive.

The empirical scale of Cofer (*Cof*), which was developed 10 years ago on a quite different college sample, is able to correctly identify 65% of the dissemblers and almost all of the authentic records at a cutting score of 20 or above. Cofer et al. (1949) reported correct identification of 86% of his dissemblers and 96% of his authentic records (phi = .641). This was accomplished by rescoring the same records from which the scale was developed. The shrinkage in percentage of correct identification in the present study would be expected on cross-validation. Decreasing the cutting score in the present sample would, of course, increase the proportion of correct identification of dissemblers at the expense of increased false positives—which could be well afforded. Gough's (1950) earlier observation that Cofer's scale "gives promise of being a very interesting addition to the pool of MMPI keys" (p. 409) is well borne out by the present results.

Table 2

Screening Efficiency of Dissimulation Measures

Scale	Cutting Score	Simulated Records ($N = 250$) Prop. Called Authen.	Simulated Records ($N = 250$) Prop. Called Simul.	Authentic Records ($N = 190$) Prop. Called Authen.	Authentic Records ($N = 190$) Prop. Called Simul.	Phi Coeff.
Sd	21 or above	.26	.74	.98	.02	.721
Cof	20 or above	.35	.65	.96	.04	.619
L	6 or above	.38	.62	.91	.09	.539
Ex	18 or above	.47	.53	.91	.09	.461
Sd–R	31 or above	.48	.52	.86	.14	.395
Sd–A	32 or above	.37	.63	.72	.28	.386
En	36 or above	.54	.46	.85	.15	.330
K	22 or above	.71	.29	.89	.11	.217

In accord with expectations, the *L* scale is indicative of dissimulation. The extent to which identification of simulated protocols could be improved is limited by the low cutoff in this 15-item scale. The rational social desirability scales of Hanley (*Ex*) and Wiggins and Rumrill (*Sd-R, Sd-A*) appear to identify dissimulated records at slightly better than chance level while minimizing the proportions of false positives rather successfully. Lower cutting scores might be of some practical value in these scales.

The rational scale of Edwards (*En*) which is probably the most widely used of current measures of "social desirability" misclassifies more dissemblers than it correctly identifies. Considering that the dissemblers in the present study were instructed to respond in terms of "social desirability" as defined by Edwards (1957), the *En* scale is clearly not what it has been assumed to be.

The present findings, with respect to *K*, add more negative evidence concerning the meaning of *K* in populations other than the original (Hunt et al., 1948; Schmidt, 1948; Tyler & Michaelis, 1953; Welsh & Dahlstrom, 1956). It is clear that high *K*, in a college sample, is not prima facie evidence of social desirability faking. It should be noted, however, that low *K* is a fairly good indication of the absence of this response tendency. Neither of these statements contradicts current clinical practices in regard to the interpretation of *K* (Welsh & Dahlstrom, 1956).

Response Bias

The need for differentiating the different "stylistic" components that enter into inventory response measures has been emphasized by Jackson and Messick (1958) and by Webster (1958). In comparing measures of social desirability response style, it seems appropriate to estimate the extent to which other response styles may be operative in accounting for the same variance. One such response style is that of "acquiescence" which has been studied in other contexts (Jackson & Messick, 1958). Fricke's "response bias" scale (*B*) may be considered a measure of acquiescence since it consists of 63 items of high "ambiguity" that have been arbitrarily scored "true." The failure of *B* to discriminate dissemblers from nondissemblers (Table 2) suggests that it represents a stylistic component mainly irrelevant to the task of identifying social desirability dissemblers. There are, undoubtedly, factors other than "response bias" which contribute to the variance of the *B* scale, but it would seem to serve as the best available operational definition of response bias tendencies. For this reason, the eight measures of dissimulation were correlated with the "response bias" scale under *both* standard and role-playing conditions. Table 3 presents these correlations.

Table 3

Correlations of Dissimulation Measures with the Response Bias Scale Under Standard and Social Desirability Instructions[a]

	Men		Women	
	Standard ($N = 105$)	Social Desirability ($N = 144$)	Standard ($N = 106$)	Social Desirability ($N = 85$)
Sd	373	−248	103	−197
Cof	−084	−426	−172	−247
L	−445	−461	−320	−500
Ex	−432	−501	−449	−532
Sd-R	126	−305	−066	−255
Sd-A	−575	−672	−633	−634
En	−594	−607	−588	−635
K	−648	−692	−615	−662

[a] For $N = 100$; an r of .257 is significant at the .01 level.

Dissimulation scales such as *Sd, Sd-R,* and Cofer's scale (*Cof*) are relatively free from this bias, while *K* and the scales of Edwards (*En*) and Hanley (*Ex*) are more subject to its effects. The response bias properties of *K* have been noted by several authors (see Jackson & Messick, 1958), and the possibility of response bias in *En* and *Ex* was noted in an earlier paper by the present author (Wiggins & Rumrill, 1959).

In addition to the individual differences in susceptibility to response bias among the dissimulation measures, it is interesting to note the different degrees of this set that appear to operate under the two conditions of administration. In general, it appears that dissimulation instructions act to *increase* the operation of the response bias factor. This is in line with Hanley's suggestion that acquiescent individuals might have a tendency to be defensive as well (Hanley, 1957).

Intercorrelations Among Dissimulation Measures

Table 4 shows the intercorrelations among dissimulation measures for the 105 men under standard instructions (lower left) and the 144 men under experimental instructions (upper right). Dissimulation scales which are slightly or insignificantly related under standard conditions become highly interrelated under instructions to answer in terms of social desirability. It might be hypothesized that a dissimulation set increases both response bias (denial) and social desirability tendencies. Since all of the scales appear to measure both of these tendencies in varying proportions, the effect of dissimulation instructions might be to increase the intercorrelations among measures.

It thus becomes difficult to choose between the scales under dissimulation instructions. Some choice is possible under standard instructions, however. Here the fallacy of developing a rational measure based on the correlation with a presumed measure of social desirability, such as *K*, becomes apparent. In general, the higher the correlation with *K*, the less empirically effective is the measure of social desirability. This could be due to the fact that *K* contributes response bias and other variance irrelevant to the identification of social desirability dissemblers (Hanley, 1957). The clear exception to this trend is found in *Sd-R*, which is not an effective empirical measure of social desirability but which is so free of response bias (Table 3) as to be virtually unrelated to *K* under standard instructions.

Table 4

Intercorrelations Among Dissimulation Measures for 105 Control Men (lower left) and 144 Experimental Men (upper right)[a]

	Sd	*Cof*	*L*	*Ex*	*Sd–R*	*Sd–A*	*En*	*K*
Sd		880	801	783	720	596	590	665
Cof	450		797	775	756	755	711	783
L	−034	353		798	582	601	572	745
Ex	115	488	359		646	702	714	848
Sd–R	348	351	090	321		629	633	606
Sd–A	045	416	253	601	253		928	818
En	062	428	282	725	286	803		778
K	−014	385	350	696	110	659	770	

[a] For $N = 125$; an *r* of .231 is significant at the .01 level.

The scale which is most similar to *Sd*, under standard instructions, is the scale of Cofer (*Cof*). This similarity is not surprising considering the manner in which the scales were developed. Cofer's scale is based on items which were changed under fake-good instructions, with the additional property that they were also left unchanged under fake-bad instructions. The development of *Sd* was, in part, a cross-validation of Cofer's procedure and, as might be expected, 14 of Cofer's items appear in the *Sd* scale. The two scales appear to be equally free of response bias (Table 3) and almost equivalent in their ability to screen social desirability dissemblers (Table 2). It appears that *Sd* and *Cof* are the best available experimental scales for investigating social desirability response tendencies in the MMPI.

Correlations between Hanley's scale (*Ex*) and *En*, *K*, *L*, and *Cof* are similar to those reported by Hanley (1957). It should be noted that, on the basis of these correlations, Hanley felt that *Cof* was of little value compared with *K*, *En*, and *Ex* (p. 394). Edwards' scale (*En*) is seen to be highly similar to *Sd-A*, as was suggested in an earlier paper (Wiggins & Rumrill, 1959). It is also substantially related to *K*, against which it was partially "validated."

Social Desirability in the MMPI

Studies by Fordyce (1956), Hanley (1957), and Edwards (1957) have done much toward creating the impression that the MMPI clinical scales are heavily saturated with social desirability response style. Such "social desirability" is a matter of definition, however, and these studies have rested on the correlations of *En* and *Ex* with the clinical scales as evidence of the susceptibility of the clinical scales to this bias. The correlations in Table 5 are especially relevant to this line of reasoning. The columns represent the dissimulation measures in *decreasing* order of their effectiveness in identifying known (instructed) social desirability dissemblers. The rows represent the MMPI scales, and the data are those

Table 5

Correlations of Dissimulation Measures and MMPI Clinical Scales for 105 Control Men Under Standard Instructions[a]

Clinical Scales	Dissimulation Measures							
	Sd	*Cof*	*L*	*Ex*	*Sd-R*	*Sd-A*	*En*	*K*
Hs	−019	−190	001	−388	−266	−496	−563	−347
D	−327	−388	085	−412	−386	−393	−434	−150
Hy	−090	−068	203	140	−169	−034	−024	348
Pd	−061	−399	−102	−454	−242	−486	−528	−349
Mf	−309	−510	−046	−312	−292	−366	−318	−177
Pa	010	−134	052	016	−122	−287	−171	034
Pt	−103	−537	−373	−698	−373	−807	−844	−751
Sc	−091	−503	−262	−616	−350	−753	−803	−638
Ma	170	−125	−199	−399	−028	−394	−476	−581
Si	−334	−303	−123	−552	−463	−575	−632	−556
A	−049	−523	−326	−666	−242	−894	−890	−781
R	−299	043	366	271	−516	275	331	573

[a] For $N = 100$; an r of .257 is significant at the .01 level.

of the 105 control men who represent a standard MMPI group comparable to those employed by others.

Table 5 shows a trend for the more accurate empirical measures of social desirability (*Sd, Cof, L*) to be only slightly correlated with the clinical scales, and for the less accurate measures (*Sd-A, En, K*) to be more highly correlated with the clinical scales. A convincing case can be made for the influence of "social desirability" in the MMPI by using the Edwards scale (*En*) as a measure of this response style. However, if one were to employ an actual empirical measure of this tendency, such as *Sd*, the case is considerably weakened. Several of the correlations of *Sd* and *Cof* with the clinical scales suggest that social desirability is indeed a factor to be reckoned with in normal populations—even though its influence has been exaggerated in previous studies.

It is also apparent, from Tables 4 and 5, that scales such as *En, Sd-A,* and *K* share a common element that runs throughout the MMPI clinical scales. Table 3 strongly suggests that this element is, in part, some form of "response bias." There are now a considerable number of studies (see Jackson & Messick, 1958) which suggest that this "response bias" is related to the primary factors measured by the MMPI and that it may well reflect *valid* (criterion-relevant) variance rather than "error" in assessment. In any event, it seems clear from the present study that the various "stylistic" components of inventory responses must be carefully separated and that recourse to empirical procedures is a requisite to eventual clarification of their role in personality inventories, such as the MMPI.

Summary

The empirical validity of several existing measures of social desirability in the MMPI was investigated by the method of contrasted groups. A group of 250 college men and women were instructed to answer the MMPI in terms of the social desirability of the items. A different group of 190 college men and women received standard MMPI instructions.

Although two empirical measures of social desirability were relatively effective in identifying dissemblers, certain rational social desirability scales, and especially that of Edwards (SD), were found to be poor predictors of the criterion. The rational social desirability scales which were ineffective in predicting dissimulation were found to be substantially correlated with an independent measure of response bias. It was suggested that previous reports of the influence of social desirability on the MMPI clinical scales would be more appropriately considered as studies of response bias. It was emphasized that these

two "stylistic" components should be differentiated and that empirical methods are the methods of choice in such an enterprise.

Received August 28, 1958.

REFERENCES

BENTON, A. L. The MMPI in clinical practice. *J. nerv. ment. Dis.*, 1945, **102**, 416–420.

COFER, C. N., CHANCE, JUNE, & JUDSON, A. J. A study of malingering on the MMPI. *J. Psychol.*, 1949, **27**, 491–499.

CORAH, N. L., FELDMAN, M. J., COHEN, I. S., GRUEN, W., MEADOW, A., & RINGWALL, E. A. Social desirability as a variable in the EPPS. *J. consult. Psychol.*, 1958, **22**, 70–72.

EDWARDS, A. L. *Edwards Personal Preference Schedule. Manual.* New York: Psychological Corp., 1954.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

EDWARDS, A. L. Personal Communication. February 20, 1958.

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

FRICKE, B. G. A response bias (B) scale for the MMPI. *J. counsel. Psychol.*, 1957, **4**, 149–153.

GOUGH, H. G. Simulated patterns on the MMPI. *J. abnorm. soc. Psychol.*, 1947, **42**, 215–225.

GOUGH, H. G. The *F-K* dissimulation index for the MMPI. *J. consult. Psychol.*, 1950, **14**, 408–413.

GOUGH, H. G. Some common misconceptions about neuroticism. *J. consult. Psychol.*, 1954, **18**, 287–292.

GRAYSON, H. M., & OLINGER, L. B. Simulation of "normalcy" by psychiatric patients on the MMPI. *J. consult. Psychol.*, 1957, **21**, 73–77.

HANLEY, C. Social desirability and responses to items from three MMPI scales: *D, Sc,* and *K. J. appl. Psychol.*, 1956, **40**, 324–328.

HANLEY, C. Deriving a measure of test-taking defensiveness. *J. consult. Psychol.*, 1957, **21**, 391–397.

HILLMER, M. L. Social desirability in a two-choice personality scale. Unpublished technical report, ONR Contract Nonr 477(08) and USPHS Grant M-743(C3), Univer. of Washington Div. of Counsel. & Testing Services, June, 1958.

HUNT, H. F. The effect of deliberate deception on MMPI performance. *J. consult. Psychol.*, 1948, **12**, 396–402.

HUNT, H. F., CARP, A., CASS, W. A., JR., WINDER, C. L., & KANTOR, R. E. A study of the differential diagnostic efficiency of the MMPI. *J. consult. Psychol.*, 1948, **12**, 331–336.

JACKSON, D. N., & MESSICK, S. Content and style in personality assessment. *Psychol. Bull.*, 1958, **55**, 243–252.

MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The MMPI: VI. The *K* scale. *J. consult. Psychol.*, 1948, **12**, 20–31.

MEEHL, P. E., & HATHAWAY, S. R. The *K* factor as a suppressor variable in the MMPI. *J. appl. Psychol.*, 1946, **30**, 525–564.

ROSEN, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.*, 1956, **52**, 151–158.

SCHMIDT, H. O. Notes on the MMPI: The *K* factor. *J. consult. Psychol.*, 1948, **12**, 337–342.

TYLER, F. T., & MICHAELIS, J. U. *K*-scores applied to MMPI scales for college women. *Educ. psychol. Measmt.*, 1953, **13**, 459–466.

VOAS, R. B. A procedure for reducing the effects of slanting questionnaire responses toward social acceptability. *USN Sch. Aviat. Med. Res. Rep.* 1956, No. NM 001 108 100, Rep. No. 17, ii-9.

WEBSTER, H. Correcting personality scales for response sets or suppression effects. *Psychol. Bull.*, 1958, **55**, 62–64.

WELSH, G. S., & DAHLSTROM, W. G. Construction. Sec. II. In *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956. Pp. 58–59.

WIGGINS, J. S., & RUMRILL, C. Social desirability in the MMPI and Welsh's factor scales *A* and *R. J. consult. Psychol.*, 1959, **23**, 100–106.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

INTELLIGENCE TEST PERFORMANCE AND THE DELAY FUNCTION OF THE EGO

GEORGE SPIVACK, MURRAY LEVINE, AND HERBERT SPRIGLE
Devereux Foundation, Devon, Pennsylvania

Many studies have shown low, but significant, correlations between Rorschach human movement (M) responses and general intelligence (Levine, Spivack, & Wight, 1959). The M response has been conceptualized as both a product and a measure of the delay function of the ego, or of inhibition ability (Singer, 1955; Singer, Wilensky, & McCraven, 1956). Since intelligence test performance may be conceptualized in terms of ego function (Fromm, Hartman, & Marschak, 1954; Rapaport, Gill, & Schafer, 1945), the question arises whether it is possible to relate more specific measures of delay, not only to general intelligence but, also, to details of intelligence test performance itself.

Our conception is similar to the position detailed by Rapaport (1951) in which the "apparatuses" of cognition develop in childhood as a function of inevitable delay in gratification, and that the further development of thinking (memory, fantasy, abstraction, problem solving, and planfulness) supports increased delay or inhibition of the expression of impulses in consideration for reality demands. With increased age, thinking becomes increasingly a substitute for direct, impulsive action and can serve as a partial discharge of tensions.

Since the many aspects of thinking are in a sense equivalent to what we test when we test intelligence, it should be possible to demonstrate a relationship between general intelligence and measures of ability to delay which bear no obvious similarity to the kinds of tasks employed in intelligence tests. Moreover, if this conception is valid, it should be possible to isolate instances in intelligence test performance wherein the ability to inhibit or delay plays an integral part in success or failure.

Method

Experimental Tasks

Time estimation. This task required that the subject (S) tell E when a 15″, 30″, and 60″ time interval had elapsed. Two estimates at each time were obtained in random order. The scores used were the sums of both estimates for each of the three time intervals. Higher scores are considered indication of better delay functioning.

This task was selected since a prior study with a similar population had shown the time sense to be related to the ability to delay impulse gratification by responding to distant incentives (Levine & Spivack, in press). Singer and Opler (1956) and Singer, Wilensky, and McCraven (1956), working with adult schizophrenics, have shown that accuracy of performance on this time estimation task is related to other measures of inhibition ability.

Stroop color-word test. This test, as adapted for the present study, required S: (a) to read aloud as quickly as possible four color names (red, green, blue, and yellow) which are printed on a white card in black ink a total of 50 times (B/W card); (b) to name as quickly as possible the color of the ink in which 50 color names were printed where the color name and the ink hue are incongruent (C/W card). Faster reading times are indicative of greater ability to control or inhibit a strong interfering habit in performing this task.

Previous investigators (Klein & Salomon, 1952; Lazarus, Baker, Broverman, & Mayer, 1957) have accepted performance on this task as a measure of an ability to regulate competing responses or to resist the effects of interfering response tendencies. This task was employed in the present study due to its conceptual similarity to a previously employed measure of "cognitive inhibition" (Meltzoff & Levine, 1954; Levine & Meltzoff, 1956).

Rorschach M tendency. The series of 26 Barron *M*-threshold inkblots (Barron, 1955) was employed, with instructions for one response per card. The score used was the total number of *M* responses produced.

The *M* tendency was employed because it has been shown to be related both theoretically and experimentally to ego delay functioning, and has been the central operational measure in much of the research in this area (Singer, 1955; Singer, Wilensky, & McCraven, 1956).

General intelligence. General intelligence was measured by the Wechsler-Bellevue intelligence scales. These measures were obtained from the clinical records of the institution and on the average the IQ tests were administered 15 months prior to the other experimental procedures.

Subjects

The population consisted of 123 emotionally disturbed adolescents in residential treatment. There were 86 males and 37 females in the group. All diagnoses were represented, except for the psychoses and chronic brain syndrome. The average age was 16 years, with a range of 12 to 19. The mean IQ was 101, with a range of 70 to 140. The distribution of IQs was essentially normal, and the standard deviation was 14.6.

Procedure

Each *S* was seen individually and was presented with the battery of tests as "an attempt to develop some new tests." The *S*s were assured that the present test results were purely of an experimental nature and would not be entered into their clinical records. The *S*s were generally cooperative, and the tasks seemed readily understandable.

Results

The intercorrelations of the experimental measures are presented in Table 1. The significant correlations between all three measures and IQ support the hypothesis that measures of ego delay are related to general intelligence.

In order to obtain a clearer picture of the meaningfulness of these correlations in terms of IQ points, the distributions of each of the three delay measures were split at the median, and mean IQs for the upper and lower half of each were compared. Only the distribution of the 60″ time estimations was employed, primarily because it correlated best with the other variables. The *t* tests for IQ differences in the three measures considered individually are all significant, as would be expected from the significant correlations. Based on any single measure, the good and poor "inhibitors" differ on the average from five to eight IQ points. When we selected those *S*s who were good and poor *on all three delay measures*, we found a mean difference between groups of 17.2 IQ points. The difference is significant, by *t* test, well below the usually accepted probability levels.

Table 1

Intercorrelations[a] Among IQ and the Measures of Ego Delay

Variable	1	2	3	4	5	6
1. Time Estimation 15″		.86	.72	−.18	−.12	.01
2. Time Estimation 30″			.88	−.28	−.04	.22
3. Time Estimation 60″				−.45	.07	.32
4. Stroop C/W[b] reading					−.27	−.53
5. Barron *M* total						.25
6. Wechsler IQ						

[a] For an *N* of 123, correlations of .19 are significant at the .05 level (two-tail test).

[b] All correlations involving C/W are partial *r*s, in which reading time for the black and white card was partialled out to control for general reading speed.

Since the tasks were taken as different measures of the delay function of the ego, the interrelationship among them becomes important. It can be noted in Table 1 that the intercorrelations are significant, with the exception of those between time estimations and *M*.

In order to relate the three ego delay measures to intelligence test performance itself, we examined the relationship between our three measures and scores on the digit span subtest of the Wechsler. This particular subtest was chosen on the basis of Wechsler's discussion of it as a test of concentration and ability to attend (Wechsler, 1944, p. 84), and our own evaluation of this task as requiring in part the ability to control or inhibit competing or extraneous thoughts and stimuli. Digit span total raw scores were related to each of the three measures of ego delay, controlling for level of general intelligence through the use of the roving median chi-square test (Cronbach, 1949). The relationship between digit span and the C/W task approached significance ($\chi^2 = 3.5$; $p = .06$) as did the relationship with time estimation ($\chi^2 = 3.0$; $p = .08$). The relationship between digit span and *M* was not significant.

Discussion

These findings provide further support for the hypothesis that measures of an ego delay function are related to general intelligence. While the correlations between IQ and the specific measures tend to be low, when we select individuals who are good and poor inhibitors consistently, the mean IQ difference (17.2 points) is rather substantial. It would appear that as we develop more reliable and more precisely defined measures of delay functioning, it should be possible to increase the correlation substantially. The low correlations among the different measures, and the fact that time estimation and *M* do not correlate, suggest that there may very well be more than one important style of control (Singer, Wilensky, & McCraven, 1956). The fact that *M* does not correlate with time estimation is consistent with other recent findings which suggest that the *M* response in childhood and adolescence does not have the same interpretive value as in adults (Levine, Spivack, & Wight, 1959; Litwin, 1957). One reservation should be kept in mind in that the present population was an emotionally disturbed one. Impulsivity is characteristic of many of the adolescents we used, and the possibility exists that in a normal group the relationship under investigation might differ either in degree or kind. It is also possible that the same study in a normal group would require more sensitive measures, since impulsivity might manifest itself more subtly and less pervasively in normals.

The finding with digit span indicates that in adolescents not only are delay mechanisms related to general level of intelligence, but that the specific operation of such mechanisms may be identified in the details of intelligence test performance itself. This finding, along with those of prior studies with adults (Levine, Glass, & Meltzoff, 1957; Levine, Spivack, & Wight, 1959) indicate that it may eventually be feasible to discuss intelligent "behavior" within the general framework of ego psychology. The task remains to explore further various relationships between measures of control and qualitative aspects of intelligence test performance. An area worth exploring is that of the nonintellective aspects of intelligence test performance. An individual with poor delaying ability may not persevere when the difficulty of the task arouses anxiety or it becomes apparent that a quick solution is not forthcoming. He might characteristically respond with the first thought that comes to mind, whereas holding himself in check and reflecting a moment might permit him to develop a response which would score two points instead of one. The possibility also exists that in individuals with good delay functioning, a positive value is placed on thinking and "intellect"—thinking functioning for such individuals as both a direct and indirect gratification. For such individuals, manipulating abstract concepts, developing vocabulary, and the idea of finding the answer may be gratifying. Such qualities are certainly assets in taking an intelligence test. A more detailed analysis of the attitudes and approach of the subject throughout the test might reveal more clearly the relationships between personality and intelligence.

Once such relationships have been clearly shown, it may be possible to develop a more unified theory in which intelligence and personality variables may be discussed in the same terms.

Summary

The purpose of the present study was to demonstrate a relationship between three measures of ego delay function, general intellectual level, and performance on the digit span subtest of the Wechsler-Bellevue (1944). The results support the hypothesis that ego delay function, as measured by M, time estimation, and the Stroop test, is correlated with IQ. Performance on the digit span subtest was shown to be related to time estimation and to the Stroop (1935) test independently of general intellectual level. The results suggest the feasibility of treating intelligence test performance and personality variables within a single theoretical framework.

Received August 28, 1958.

REFERENCES

Barron, F. Thresholds for the perception of human movement in inkblots. *J. consult. Psychol.,* 1955, **19**, 33–38.

Cronbach, L. J. Statistical methods applied to Rorschach scores: A review. *Psychol. Bull.,* 1949, **46**, 393–429.

Fromm, Erika, Hartman, Lenore D., & Marschak, Marian. A contribution to a dynamic theory of intelligence testing of children. *J. clin. exp. Psychopath.,* 1954, **15**, 73–95.

Klein, G. S., & Salomon, A. Cognitive style and recognition of need. *Amer. Psychologist,* 1952, **7**, 321–322. (Abstract)

Lazarus, R. S., Baker, R. W., Broverman, D. M., & Mayer, J. Personality and psychological stress. *J. Pers.,* 1957, **25**, 559–577.

Levine, M., Glass, H., & Meltzoff, J. The inhibition process, Rorschach human movement responses and intelligence. *J. consult. Psychol.,* 1957, **21**, 41–45.

Levine, M., & Meltzoff, J. Cognitive inhibition and Rorschach human movement responses. *J. consult. Psychol.,* 1956, **20**, 119–122.

Levine, M., & Spivack, G. Incentive, time conception and self control in a group of emotionally disturbed adolescents. *J. clin. Psychol.,* in press.

Levine, M., Spivack, G., & Wight, B. The inhibition process, Rorschach human movement responses and intelligence: Some further data. *J. consult. Psychol.,* 1959, **23**, 306–312.

Litwin, Dorothy. Sensory-tonic theory as it relates to the development of inhibition and social maturity. Unpublished doctoral dissertation, New York Univer., 1957.

Meltzoff, J., & Levine, M. The relationship between motor and cognitive inhibition. *J. consult. Psychol.,* 1954, **18**, 355–358.

Rapaport, D. *Organization and pathology of thought.* New York: Columbia Univer. Press, 1951.

Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing.* Chicago: Year Book, 1945.

Singer, J. L. Delayed gratification and ego-development: Implications for clinical and experimental research. *J. consult. Psychol.,* 1955, **19**, 259–266.

Singer, J. L., & Opler, M. K. Contrasting patterns of fantasy and motility in Irish and Italian schizophrenics. *J. abnorm. soc. Psychol.,* 1956, **53**, 42–47.

Singer, J. L., Wilensky, H., & McCraven, Vivian G. Delaying capacity, fantasy and planning ability: A factorial study of some basic ego functions. *J. consult. Psychol.,* 1956, **20**, 375–383.

Stroop, J. R. Studies of interference in serial verbal reaction. *J. exp. Psychol.,* 1935, **18**, 643–661.

Wechsler, D. *The measurement of adult intelligence.* Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

ERRORS IN THE DIAGNOSIS OF MENTAL DEFICIENCY IN SCHIZOPHRENIA [1]

LOREN J. CHAPMAN
University of Chicago

AND JULIAN H. PATHMAN
Downey Veterans Administration Hospital

Since symptoms which resemble mental deficiency often result from the schizophrenic process, it is sometimes difficult to distinguish patients who are schizophrenic but not mentally defective from those who are both. The resultant errors may be deleterious to the nondefective schizophrenic, in that they inhibit appropriate therapy.

The present paper reports a study of the characteristics of a number of cases in which, after a psychiatric interview, a diagnosis of mental deficiency was erroneously made in addition to the finding of schizophrenia.

An inspection of the records at a neuropsychiatric hospital revealed that 41 out of approximately 2,400 patients bore a diagnosis of both mental deficiency and psychosis.

An attempt was made to test each of the 41 with the Wechsler-Bellevue Intelligence Scale, Form I. Twenty-nine were testable, and 17 made IQ scores of 70 or above, despite the fact that there were clear indications in most of their protocols of psychotic disruption of test performance. This group of 17 patients was chosen for further study.

The mean IQ score for the 17 patients was 85.9. The range was 71 to 103. Everyday clinical experience and considerable research indicate that chronic schizophrenia reduces functioning IQ. While it is difficult to obtain an exact estimate of the amount of reduction, reports of previous findings (Gilliland, Wittman, & Goldman, 1943; Hunt & Cofer, 1944) indicate that an estimate of 15 points is not unreasonable. It would appear, therefore, that despite their diagnosis, the 17 patients were, as a group, in the average range of intellectual endowment. The mean age of these patients was 56.9 years. Eight of them were over 59 years of age, the maximum age for which Wechsler provides IQ tables. For these eight patients, the weighted score-IQ equivalents for ages 55–59 were used. This procedure is conservative, for purposes of this study, since it slightly underestimates the IQs.

The patients had been diagnosed on the basis of a psychiatric interview soon after admission. The 17 diagnoses had been made by nine different psychiatrists.

The main purpose of this investigation was to determine how these 17 schizophrenic patients, who were misdiagnosed as mental defective as well as psychotic, differed from other patients who were diagnosed only as psychotic.

In order to make the comparison, a 26-patient control group, chosen from the same wards, was selected so that they had the same mean year of birth (1896) and mean year of admission (1940) as the experimental group. For each of these measures the variances were almost identical for the two groups.

The experimental and control groups were compared with respect to educational and occupational level, two social history variables which were thought to be of possible importance. The control group was found to have a slightly higher mean educational level (8.2 years) than the experimental group (6.7 years), but the difference was not statistically significant ($p = .36$).

The highest occupational level attained by each *S* was scored, using the median Army

---

[1] This study was supported by the Social Science Research Committee of the University of Chicago.

General Classification Test score for World War II inductees as reported by Stewart (1947). The purpose of this comparison was to determine whether a patient was likely to be diagnosed as mentally defective because his occupation was one usually held by people of low intelligence. The mean A.G.C.T. score corresponding to the occupations was 95.6 for the experimental group and 105.9 for the controls. The difference was significant at the .01 level.

The two groups were next compared on diagnosis aside from the finding of mental deficiency. Although both groups consisted entirely of patients described as functional psychotics, they were found to differ markedly in the number of patients who were labeled in terms of the classic subtypes of schizophrenia. Of the experimental group, less than one-third were diagnosed either paranoid, hebephrenic, catatonic, or simple, while all of the control patients received such a diagnosis. Over two-thirds of the experimental patients received a diagnosis of "psychosis with mental deficiency," or some similar label. In none of the cases was there mentioned a variety of psychosis other than schizophrenia. This observation raised the question of whether the experimental patients may have received their erroneous diagnosis in part because they lacked those distinctive schizophrenic symptoms which make it possible to diagnose the patient in terms of classic subtypes.

It therefore seemed desirable to compare the two groups of patients on their symptomatology at the time of intake. The psychiatrist's description of the patient in the intake interview was copied from each patient's folder. These descriptions were found to be of about equal length for the two groups, showing a mean of 33 lines for the experimental group and 32 lines for the control group. Using these descriptions, three psychologists rated each patient on the Multidimensional Scale for Rating Psychiatric Patients (Lorr, 1953), using only those 42 items which were designed for use by an interviewing psychiatrist.

The three sets of ratings were compared for interrater reliability by means of an intraclass correlation coefficient (Haggard, 1958). It was found that the ratings were reliable above chance ($p < .05$) on all except two items (Nos. 14 and 26). These two items were eliminated from further computations.

The Multidimensional scale allows the rater to mark an item as undecided. Whenever two of the three raters marked an item "undecided" for a given patient the item was dropped for that patient, but if only one judge marked it undecided, the ratings of the other two judges were used, and a mean rating was obtained for each item on each *S*.

The items of the Multidimensional scale are grouped by the test authors into 12 symptom clusters or "factors," 8 of which are composed primarily of items from the psychiatric rating portion of the scale. The two groups of patients were compared on their mean pathology rating on each of these eight clusters. The two clusters labeled by the test authors as conceptual disorganization and perceptual distortion were found to discriminate the two groups of patients, and both were in the predicted direction. That is, the experimental group was rated as significantly less pathological than the control group, the difference being significant at the .05 level, computed by a two-tailed Mann-Whitney test. The symptom clusters on which the two groups did not differ significantly were those labeled paranoid projection, melancholy agitation, motor disturbances, hysteric conversion, self-depreciation vs. grandiose expansiveness, and retarded depression vs. manic excitement.

The two symptom clusters which distinguished the groups might be thought to be those which interviewers would consider most clearly indicative of psychosis. To verify this impression, the eight cluster labels, together with the designations of their several respective item dimensions, were submitted to 13 psychiatrists to rank order on the extent to which the symptom clusters are important for distinguishing psychosis from nonpsychosis.

The two predicted symptom clusters were rated as the most important. The mean rankings for perceptual distortion and conceptual disorganization were 1.8 and 2.5, respectively, while the mean rankings of the other six factors were, in the order listed above, 4.4, 6.1, 3.6, 7.5, 5.7, and 4.5.

It appears, then, that those patients bearing the additional diagnosis of mental deficiency showed fewer symptoms that are usually thought of as identifiably psychotic. Although the interviewing psychiatrists saw sufficient signs of psychosis to diagnose the patients as such, they evidently responded to the paucity of distinctively psychotic symptoms by adding the diagnosis of mental deficiency.

Summary

Seventeen schizophrenic patients who, on the basis of an interview, were misdiagnosed as mentally defective in addition to their diagnosis as psychotic were compared with 26 controls who were diagnosed only as schizophrenic. The experimental group was found to have attained a lower occupational level than the controls, and was reported to have exhibited less distinctively psychotic symptoms in diagnostic interviews.

Received September 2, 1958.

REFERENCES

Gilliland, A. R., Wittman, Phyllis, & Goldman, M. Patterns and scatter of mental abilities in various psychoses. *J. gen. Psychol.*, 1943, **29**, 251–260.

Haggard, E. A. *Intraclass correlation.* New York: Dryden, 1958.

Hunt, J. McV., & Cofer, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* Vol. II. New York: Ronald, 1944. Pp. 971–1032.

Lorr, M. Multidimensional scale for rating psychiatric patients, hospital form. *V. A. tech. Bull.* Washington, D. C., 1953, No. TB 10-507.

Stewart, Naomi. A.G.C.T. scores of army personnel grouped by occupation. *Occupations,* 1947, **26**, 5–41.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

EXPRESSED ACCEPTANCE OF SELF AND OTHERS BY PSYCHOTHERAPISTS[1, 2]

JULIAN W. STREITFELD
Teachers College, Columbia University

The major focus of research in psychotherapy has been on the evaluation of its effects. Comparatively little has been done to study the qualities of psychotherapists which may contribute to the success of psychotherapy. Most theoreticians and therapists agree that "the psychology of the psychologist . . . enters into the determination of the therapeutic product" (Rogers, 1949; Shoben, 1949, p. 367). Research has shown that people rated as better therapists are seen by their peers, colleagues, and supervisors to be happier, less anxious, better adjusted, and more likeable persons than are those rated as poorer therapists (Bandura, 1956; Kelley & Fiske, 1951; Luborsky, 1952).

Almost all schools of psychotherapy accept the thesis that good psychotherapy involves a good interpersonal relationship in which the patient or client can feel accepted, and in which the therapist must be an accepting person. Research findings have tended to support the hypothesis that better, more expert therapists are more accepting than are poorer therapists. This conclusion has been deduced from indirect evidence (Fiedler, 1950a; Fiedler, 1950b; Strupp, 1957), or from colleagues', peers', or supervisors' evaluations of training therapists (Kelley & Fiske, 1951; Luborsky, 1952). No investigations, however, have directly tested the widely held proposition that better therapists are more accepting persons.

Since theory has suggested (Fromm-Reichman, 1949; Rogers, 1951) and research has demonstrated (Berger, 1952; Sheerer, 1949; Stock, 1949) that a person's ability to accept others is positively related to his feelings of self-acceptance, these attitudes might well be investigated together.

The purpose of the present study was to make a direct test of the hypothesis that better psychotherapists are more accepting of others and more self-accepting than are poorer psychotherapists. The measures of the psychotherapists' attitudes are in terms of expressed acceptance of others and expressed self-acceptance.

Hypotheses

It was hypothesized that therapists who are more accepting of the self and others are better therapists, as rated by their supervisors. They also rate their psychotherapeutic ability more accurately, that is, in better agreement with their supervisors' ratings. It was also hypothesized that, since the more accepting therapists rate themselves more accurately, the relationship between the self-ratings and the degree of acceptance is greater for the more accepting therapists than for the less accepting therapists.

Corollary hypotheses were that the supervisors and the subjects, respectively, are internally consistent. For both the supervisors and the subjects, ratings of case movement and of psychotherapeutic ability are positively related. A final corollary hypothesis was that expressed acceptance of self is positively related to expressed acceptance of others.

---

[1] This paper is based upon a dissertation submitted in candidacy for the degree of doctor of philosophy at Teachers College, Columbia University. Appreciation is expressed to the members of the dissertation committee, Joel R. Davitz, Laurance F. Shaffer, and Rosalea A. Schonbar. The original report (Streitfeld, 1958) may be consulted for more detailed treatment of the rationale, methods, and results.

[2] Presented at the American Psychological Association meeting at Washington, August 28, 1958.

Subjects

Two groups of *S*s were used in the experiment. The *S*s were third- and fourth-year doctoral candidates in clinical, counseling, and school psychology, most of whom had completed a year's internship.

Group 1 consisted of 28 graduate students who were enrolled in a practicum in psychological counseling at Teachers College, Columbia University, during the academic year 1956–1957. Group 2 consisted of 51 graduate students or graduates who had been in the practicum during the academic years 1953–1954, 1954–1955, 1955–1956, and who responded to the request for data. Of 76 former students from whom the data was requested, 51 or 67% responded.

The practicum runs for an entire academic year. Students see clients in a clinic, integral with the department, under intensive supervision of professional members of the staff, in supervisory groups of three or four.

Procedure

The procedure involved the use of scales of expressed acceptance of others and expressed self-acceptance which were constructed by Berger (1952) and rating scales of psychotherapeutic ability and of case movement devised specifically for the present experiment.

Measures Obtained

Six measures were obtained from the *S*s and the supervisors.

The two measures obtained from the supervisors were their ratings of the psychotherapeutic ability of the *S*s in their past and current supervisory groups, and their ratings of the therapeutic movement of each client seen in counseling by each *S* in their current supervisory groups, Group 1. A total of 91 clients was seen by all of the *S*s. The *S*s saw from two to five clients each.

Three measures were obtained from all of the *S*s; namely, ratings of their own psychotherapeutic ability and expressed acceptance of others (AO) and expressed self-acceptance (SA), as measured by Berger's (1952) scales. In addition, the Group 1 *S*s rated the therapeutic movement of each of their clients.

Ratings of the movement of the counseling cases were obtained for Group 1 *S*s only, since their clients had been most recently seen and were, therefore, more likely to be rated accurately by both the *S*s and the supervisors.

Rating Scale of Psychotherapeutic Ability

The rating scale of over-all psychotherapeutic ability was as follows:

6—One of the best	07%
5—Considerably above average	18%
4—Somewhat above average	25%
3—Somewhat below average	25%
2—Considerably below average	18%
1—One of the poorest	07%

The supervisors were asked to rate the *S*s in the six categories, using as a reference group all students they had known at a similar level of training. Further, they were asked to select three individual students whom they considered "best," "average," and "poorest" in psychotherapeutic ability, and to think of these three students as specific reference points in making the ratings.

The *S*s were asked to rate themselves in the above categories, using as a reference group all the students they had known at a similar level of training.

Rating Scale for the Movement of Cases

The rating scale for the movement of each case contained two steps, as shown:

Step I	Step II	
3—Above average	6—Great progress	07%
	5—Considerable progress	18%
2—Average	4—Above average	25%
	3—Below average	25%
1—Below average	2—Very little progress	18%
	1—No or almost no progress	07%

The supervisors and *S*s were individually asked to rate the movements of the cases first in the categories under Step I, and then with the scores under Step II. They were asked to use as a reference group the movement of all clients or patients in counseling or psychotherapy with similarly trained students they have known.

Berger's Scales

Data on the development, reliability, and validity of Berger's scale for expressed acceptance of others and expressed self-acceptance were published by him (1952). Since the scales reflect expressed feelings of acceptance, a basic and realistic acceptance of self and others may not necessarily be revealed. Berger expressed the belief that "Immature and unrealistic persons may score *very* high on SA & AO."[3]

In using the Berger scales in the present study, it was assumed that most graduate students and graduates of a clinical, counseling, or school psychology training program would be sufficiently mature, realistic, and secure so that a basic and realistic acceptance of the self and others may be inferred from the scores they achieve on the SA and AO scales.

Since Berger's scales were validated under conditions of anonymity of *S*s, a procedure was devised, in the administration of the scales for the present study, to maintain similar conditions. The responses and ratings of the *S*s were kept unidentifiable to the supervisors and the experimenter. The *S*s were so informed in advance.

To determine the variability and reliability of the Berger scales with the population used in the present study, a pilot study was conducted in which the scales were administered to the Group 1 *S*s and the results compared with those obtained by Berger (1952). A comparison of the pilot study group with Berger's seven groups indicated an adequate spread of scores on both Berger scales. Split-half reliabilities of .79 and .99 were found for the AO and SA scales, respectively, indicating adequate reliability for use in the present study.

Results and Conclusions

All the hypotheses were evaluated with Pearson product-moment correlations, and tested for significance with the *t* test, except where otherwise indicated. In each case the probability was estimated on the basis of a two-tailed test. The mean, standard deviation, and range for each scale are presented in Table 1.

Table 1

Means, Standard Deviations, and Ranges of Scales Used in the Present Study

Scale	*N*	Mean	*SD*	Range
Expressed acceptance of others	79	115.76	9.66	95–132
Expressed self-acceptance	79	145.35	16.87	93–174
Ratings of psychotherapeutic ability by supervisors	79	4.37	.128	1–6
Ratings of psychotherapeutic ability by *S*s	79	4.57	.86	2–6
Ratings of psychotherapeutic movement of clients by supervisors	91	3.82	1.31	1–6
Ratings of psychotherapeutic movement of clients by *S*s	91	3.64	1.65	1–6

Main Hypotheses

The hypothesis that therapists who are more accepting of the self and others are better therapists, as rated by their supervisors, was tested by correlating supervisors' ratings of *S*s' psychotherapeutic ability with *S*s' scores on Berger's SA and AO scales. Nonsignificant correlations of −.02 and .09, respectively, were obtained, which failed to support the hypothesis. Correlations of supervisors' ratings of *S*s' psychotherapeutic ability with *S*s' scores on Berger's scales were also obtained for the Group 1 and Group 2 *S*s separately. None of the correlations was significant.

The hypothesis that the more accepting therapists rate their psychotherapeutic ability more accurately, that is, in better agreement with their supervisors' ratings, was tested by correlating *S*s' scores on the SA and AO scales with the differences between the *S*s' self-ratings and the supervisors' ratings of them. A significant negative correlation would confirm the hypothesis. Two measures of the differences between the supervisors' ratings of *S*s and *S*s' self-ratings were obtained. One measure was the differences without signs, the absolute differences; the other measure was the algebraic differences. To avoid the use of negative numbers, a constant, 10, was added to each algebraic difference.

---

[3] E. M. Berger. Personal communication. February 27, 1957.

None of the correlations between the scores on the acceptance scales and the differences, both absolute and algebraic, between supervisors' and self-ratings were significantly different from zero. Nonsignificant correlations also resulted for *S*s who scored in the upper and lower 50% on the AO scale and on the SA scale. As the *S*s' scores on Berger's scales are independent of the differences between *S*s' self-ratings and supervisors' ratings of *S*s' psychotherapeutic ability, the hypothesis was not supported.

The hypothesis that the relationship between the self-ratings and the degree of acceptance is greater for the more accepting therapists than for the less accepting therapists was tested by correlating the *S*s' scores on Berger's scales with *S*s' ratings of their own psychotherapeutic ability for *S*s scoring above and *S*s scoring below the fiftieth percentile on each of the Berger scales. The significance of the differences between the correlations of the upper and the lower 50% on each of the Berger scales was tested by the transformation of the correlations into *z* scores.

Table 2

Correlations of Subjects' Scores on Berger's Scales with Self-Ratings by Subjects

Berger's Scales	*N*	Correlations with Self-Ratings	Differences Between Correlations of High and Low 50%
Expressed acceptance of others			
High 50%	39	.21	.48*
Low 50%	39	−.27	
All *S*s	79	.06	
Expressed self-acceptance			
High 50%	39	.01	.04
Low 50%	39	−.03	
All *S*s	79	.28*	

* Significant at .05 level.

Table 3

Correlations of Ratings of Case Movements by Supervisors and Subjects, with Ratings of Subjects ($N = 28$) by Supervisors and Subjects

	Ratings of *S*s	
Ratings of Case Movements	By Supervisors	By *S*s
By supervisors		
Highest rating per *S*	.60**	.19
Lowest rating per *S*	.65**	−.09
Mean rating per *S*	.79**	.19
By subjects		
Highest rating per *S*	.06	.51**
Lowest rating per *S*	.36	.03
Mean rating per *S*	.42*	.41*

** Significant at .01 level.
* Significant at .05 level.

As shown in Table 2, the correlation of *S*s' self-ratings with *S*s' scores on Berger's AO scale is positive for *S*s scoring above the fiftieth percentile on the AO scale, and negative for the *S*s scoring below the fiftieth percentile. Though the difference between the two correlations is significant at the .05 level, neither correlation, by itself, is significantly different from zero. No significant difference was found between the correlations of *S*s' self-ratings and their scores on Berger's SA scale for *S*s scoring above the fiftieth percentile and *S*s scoring at and below the fiftieth percentile on the SA scale. While there is a small but significant relationship of *S*s' self-ratings and their scores on the SA scale for all *S*s, there is no such relationship for *S*s scoring at and below the fiftieth percentile, or for *S*s scoring above the fiftieth percentile on the SA scale. The hypothesis was partially supported.

Corollary Hypotheses

The corollary hypotheses, that for both the supervisors and the *S*s, the ratings of case movement are positively related to ratings of therapeutic ability, were tested by correlating the highest, lowest, and mean ratings of the therapeutic movement of the clients seen by the Group 1 *S*s with the respective supervisors' and *S*s' ratings of the *S*s' psychotherapeutic ability. As indicated in Table 3, with the sole exception of the lowest ratings of case movement by the *S*s correlated with the *S*s' self-ratings, the hypotheses were supported by correlations significant at the .01 and .05 levels.

The final corollary hypothesis, stating that expressed acceptance of self is positively related to expressed acceptance of others, was tested and supported by a correlation of .53 between *S*s' scores on the two Berger scales, which is significant at the .01 level.

Additional Findings

As shown in Table 3, except for the positive correlation, significant at the .05 level, of the mean rating of case movement by *S*s and supervisors' ratings of *S*s' psychotherapeutic ability, ratings of case movements by supervisors are independent of *S*s' self-ratings of psychotherapeutic ability, and *S*s' ratings of case movements are independent of supervisors' ratings of *S*s' psychotherapeutic ability.

There is a positive correlation of .56, significant at the .01 level, of supervisors' and *S*s' ratings of case movement.

Table 4 shows that there is a positive relationship, significant at the .01 level, between supervisors' ratings of *S*s' psychotherapeutic ability and *S*s' self-ratings. The positive relationship exists for *S*s scoring below the fiftieth percentile on the AO scale and for *S*s scoring below the fiftieth percentile on the SA scale, but not for the *S*s scoring above the fiftieth percentile on either of the scales. The difference between the correlations of low and high *S*s on both scales, as tested by the transformation of the correlations into *z* scores, was not significantly different from zero.

Table 4

Correlations of Supervisors' Ratings of Subjects with Self-Ratings by Subjects

Subjects Rated by Supervisors	*N*	Correlations with Self-Ratings by *S*s	Differences Between Correlations of High and Low 50%
High 50% of expressed acceptance of others	39	.19	.35
Low 50% of expressed acceptance of others	39	.54**	
High 50% of expressed self-acceptance	39	.26	.24
Low 50% of expressed self-acceptance	39	.50**	
All *S*s	79	.40**	

** Significant at .01 level.

Further analysis of the data revealed that of the 39 *S*s who scored at and below the fiftieth percentile on the AO scale, 28 *S*s also scored at and below the fiftieth percentile on the SA scale. The correlations of the supervisors' ratings of *S*s' psychotherapeutic ability with *S*s' self-ratings were .50 for the 28 *S*s who scored at and below the fiftieth percentile on both scales, .13 for the 11 *S*s who scored low on only the AO scale, and .74 for the 11 *S*s who scored low on only the SA scale. The correlation for the *S*s who scored low on only the AO scale is not significantly different from zero. The other correlations are significant at the .01 level.

Discussion

The confirmation of the corollary hypotheses indicated at least an internal consistency by the supervisors and by the students, and some measure of concurrent validity of the instruments used in the study.

The therapists' self-ratings of their psychotherapeutic ability are most closely related to their highest ratings of the psychotherapeutic movement of their clients. They are also related to a lesser extent to their mean ratings of case movement, and are not at all related to their lowest ratings of client psychotherapeutic movement. The findings strongly suggest that the therapists tended to perceive their own psychotherapeutic ability more in terms of their psychotherapeutic successes than in terms of their comparative failures.

Expressed Acceptance of Self and Others, and Psychotherapeutic Ability

The results of the present study failed to support the hypothesis that better psychotherapists are more accepting of others or more self-accepting than are poorer psychotherapists. They indicate that, among selected graduate students in clinical, counseling, and school psychology training programs, there is no relationship between expressed acceptance of self or others and judged psychotherapeutic ability.

The one positive finding was that the correlation of therapists' self-ratings of their therapeutic ability with their expressed acceptance of others was positive for the therapists expressing more acceptance of others, and negative for the therapists having less expressed acceptance of others, the difference being significant at the .05 level. The finding may be due to chance since neither of the correlations was, in itself, significantly different from zero.

An additional finding suggesting further examination is that therapists having comparatively low expressed acceptance of self and others and those having only comparatively low expressed self-acceptance perceived their psychotherapeutic ability more as their supervisors evaluated them than did the therapists with greater expressed acceptance of self and others and even those with only lower expressed acceptance of others. The findings suggest that the greater degree of similarity may be related to low self-acceptance rather than to high self-acceptance as postulated. Perhaps the less self-accepting therapists are less secure and do not form as independent judgments as do the more self-accepting therapists. They might seek evaluation from their supervisors more frequently and be more sensitive to their supervisors' opinions than the more self-accepting therapists.

Another possible explanation for the findings that the low scorers on AO and SA rate their therapeutic ability more as their supervisors rate them than do the high scorers is that Berger's idea that "Immature and unrealistic persons may score *very* high on SA and AO" is correct and applicable to the *S*s in the present study contrary to the assumption made in the study. If so, the high scorers on the "acceptance" scales would be less realistic and therefore less "accurate" in their self-ratings. Similarly, Berger's idea may account for the positive correlation of self-ratings and AO scores for the high AO scorers, and the negative correlation for the low AO scorers. Unrealistically high self-ratings by the high scorers might account for the positive correlation. To the extent that the high scorers on Berger's scales are unrealistic, the conclusions based on the present study would be limited.

Implications

The findings suggest that, contrary to the common opinions among psychotherapists and theoreticians, better therapists are not necessarily more accepting people. If it is true that acceptance of the patient or client is necessary to good therapy, it may be that better psychotherapists are capable of being accepting of their patients or clients during the psychotherapeutic hours. They would not necessarily have to be accepting of all people, or people in general, all of the time. Fiedler and Senior (1952) suggest the possibility that therapists are able "to 'soft pedal' and/or disguise those aspects of their personalities which would be distressing to the patients, and emphasize, instead, those aspects of their personalities which are reassuring and inspiring" (p. 452). Rogers (1957) wrote that "It is not necessary . . . that the therapist be a paragon who exhibits this degree of integration, of wholeness, in every aspect of his life. It is sufficient that he is accurately himself in this hour of this relationship . . ." (p. 97). Thus it may be necessary, in the investigation of the contribution of the therapist's personality to good psychotherapy, to study the therapist's personality as it is manifested in the therapeutic situation. Notwithstanding this possible modification, this study does not support the hypothesis that general acceptance of self or others is related to psychotherapeutic ability.

Summary

A direct test was made of the hypothesis that better psychotherapists are more accepting of others, and more self-accepting, than are poorer psychotherapists. This belief is held by most psychotherapists and schools of psychotherapy.

The 79 subjects were current and former students in the doctoral programs in clinical, counseling, and school psychology at Teachers College, Columbia University, who were taking, or had taken in the previous three years, an intensive required practicum in psychological counseling.

The criterion of psychotherapeutic ability was the ratings of the subjects' supervisors in the practicum. Each supervisor supervised three or four students per semester. Therapeutic movement of each subject's clients was also rated. The acceptance variables were measured by Berger's scales of Expressed Self-Acceptance and Expressed Acceptance of Others.

The results failed to support the hypothesis. No relationship was found between psychotherapeutic ability, as rated by supervisors, and expressed acceptance of self or expressed acceptance of others. Ratings of therapeutic ability were found to be related to ratings of therapeutic movement of clients, and expressed acceptance of self was found to be related to expressed acceptance of others.

Though there are limitations of the study in terms of the scales and subjects used, the results suggest that better therapists are not necessarily more accepting persons. If acceptance is a necessary ingredient in the therapeutic process, it may be that better therapists are more accepting in the specific situations of therapy.

Received September 8, 1958.

REFERENCES

Bandura, A. Psychotherapist's anxiety level, self-insight, and psychotherapeutic competence. *J. abnorm. soc. Psychol.*, 1956, **52**, 333–337.

Berger, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.*, 1952, **47**, 778–782.

Fiedler, F. E. A comparison of therapeutic relationships in psychoanalytic, nondirective and Adlerian therapy. *J. consult. Psychol.*, 1950, **14**, 436–445. (a)

Fiedler, F. E. The concept of an ideal therapeutic relationship. *J. consult. Psychol.*, 1950, **14**, 239–245. (b)

Fiedler, F. E., & Senior, Kate. An exploratory study of unconscious feeling reactions in fifteen patient therapist pairs. *J. abnorm. soc. Psychol.*, 1952, **47**, 446–453.

Fromm-Reichman, Frieda. Remarks on the philosophy of mental disorders. In P. Mullahy (Ed.), *A study of interpersonal relations*. New York: Hermitage Press, 1949. Pp. 162–191.

Kelley, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. Michigan Press, 1951.

Luborsky, L. B. The personality of the psychotherapist. *Menninger Quart.*, 1952, **6** (4), 1–6.

Rogers, C. R. The attitude and orientation of the counselor in client-centered therapy. *J. consult. Psychol.*, 1949, **13**, 82–94.

Rogers, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.

Rogers, C. R. The necessary and sufficient conditions of therapeutic personality change. *J. consult. Psychol.*, 1957, **21**, 95–103.

Sheerer, Elizabeth. An analysis of the relationship between acceptance of and respect for the self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169–175.

Shoben, E. J., Jr. Psychotherapy as a problem in learning theory. *Psychol. Bull.*, 1949, **46**, 366–392.

Stock, Dorothy. An investigation into the interrelations between the self concept and feelings directed toward other persons and groups. *J. consult. Psychol.*, 1949, **13**, 176–180.

Streitfeld, J. W. Expressed acceptance of self and others in psychotherapists. Unpublished doctoral dissertation, Teachers Coll., Columbia Univer., 1958.

Strupp, H. H. A multidimensional comparison of therapist activity in analytic and client-centered therapy. *J. consult. Psychol.*, 1957, **21**, 301–308.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

DISCRIMINATION OF FEMALE SCHIZOPHRENICS WITH CONFIGURAL ANALYSIS OF THE MMPI PROFILE

WILLIAM J. EICHMAN[1]

Veterans Administration Hospital, Roanoke, Virginia

Although the MMPI was originally designed to discriminate between various diagnostic categories, early research reported discouraging results in this respect. These early studies were largely confined to *t* tests between groups on individual scales. Recently, more complex approaches have been used, among which is a configurational analysis of paired scores within a particular profile. Sullivan and Welsh (1952), who developed the methodology, demonstrated some success with this method in discriminating ulcer patients from controls. Taulbee and Sisson (1957), using the same methodology, were able to discriminate schizophrenics from neurotics in a sample of male VA patients. In both studies, cross-validation showed good reliability for the signs.

The Taulbee-Sisson study with schizophrenics appears to be more important in regard to practical application of the results. Nevertheless, the use of their signs upon a sample of female VA patients yielded discouraging results and led to the present study. It is not surprising to find different results in a different hospital setting, particularly when patients of the opposite sex are used as subjects. The question remains whether these results are due to different diagnostic practices or to real differences in performance between the sexes.

In addition to the above, the Taulbee-Sisson study has limited utility. Their control group was composed of neurotic patients, making it necessary to screen out, by other means, all character and personality disorders, organics, manic-depressives, etc. This presents an awesome problem, particularly in regard to the character and personality disorders, since these groups show symptomatology which appears to be between that of the neurotic and schizophrenic groups, both qualitatively and quantitatively. The present study combines these nosological categories in the control group. Thus the problem is to develop a set of signs from the MMPI that will distinguish between two groups of female patients: (*a*) an unselected heterogeneous group of women diagnosed as neurosis, character or behavior disorder and (*b*) an unselected, heterogeneous group of women who had a primary diagnosis of schizophrenia. Considerable refinement of Sullivan-Welsh methodology can be made. Those refinements which were included in the present study are: (*a*) the use of the validating scales in the development of signs; (*b*) the use of numerical difference scores rather than the simple "+" or "−" used in the two previous investigations; (*c*) cross-validation of signs in other hospital settings in addition to the one in which the signs were developed and; (*d*) application of weights to the signs which will maximize the discrimination between groups.

A secondary problem which was investigated concerned the use of the *K* scale as a suppressor variable. Since the introduction of this scale, it has been almost routine clinical practice to add a prescribed proportion of the *K* score into the scores of five of the clinical scales before converting raw scores into standard scores. Although discrimination between diagnostic groups might be improved by such a practice, important statistical disadvantages are introduced, i.e., experimental independence is reduced. The authors (McKinley, Hathaway, & Meehl, 1948) of the *K* scale introduced these weights with the caution that other weights might be more practical in other situations. This problem has never been fully investigated. The approach used in this study was to develop signs from both the *K* weighted and the unweighted profiles, utilizing that set of signs which offered more potential.

---

[1] The author wishes to thank Hayden L. Mees, VA Hospital, American Lake, Washington, Ralph Simon, VA Hospital, Perry Point, Maryland, and John Altrocchi, Duke University, for providing protocols for cross-validation.

Method

Subjects

The *S*s used for the validation of signs included all female veterans in the Veterans Administration Hospital, Roanoke, Virginia, who had been administered the MMPI and who had received diagnoses fitting the criteria described below. Approximately half these *S*s had been administered the test on admission as part of a battery, and the test results had some influence on the diagnosis. The remaining half of the *S*s were administered the test for different purposes, e.g., prepsychotherapy evaluation, evaluation of current status, etc., and the test had no influence on diagnosis, e.g., the test may have been administered a year after diagnosis had been assigned. This latter group, however, included many chronic patients whose symptomatic picture had changed and who would not have received the same diagnosis were they to be re-evaluated.

Control Group

Thirty-three patients were found who were diagnosed neurosis, personality or character disorders. The percentage of patients in each group was: (*a*) neurosis 62%, (*b*) character or personality disorders 38%. The average age of the group was 37.0, with a standard deviation of 10.1 years.

Schizophrenic Group

Fifty-six patients were found who had a diagnosis of schizophrenia. These broke down according to the following categories: (*a*) chronic, undifferentiated, 52%; (*b*) paranoid, 35%; (*c*) catatonic, 4%; (*d*) schizo-affective, 4%; (*e*) hebephrenic, 4%; and (*f*) simple, 2%.

The average age of the group was 33.6, with a standard deviation of 6.6 years. Many of these patients showed little or no overt psychotic symptomatology. Initially, this study had been planned to use two schizophrenic groups, one with overt psychotic symptomatology and one with subtle psychotic symptomatology. The judgment of "subtle" or "overt" was made on the basis of whether one or more of the following existed at the time of the test administration: (*a*) delusions, (*b*) hallucinations, (*c*) gross confusion of thought processes, (*d*) gross loss of control over behavior, (*e*) profound withdrawal, (*f*) other bizarre behavior. Of the 56 patients who comprised the total schizophrenic group, 28 were assigned to each group. Thus half the validation group showed no immediate signs of schizophrenic psychosis at the time of administration of the MMPI. Results are reported for each of these groups separately.

Sign Analysis

Seventy-eight scores were obtained for each patient. Twelve of these were the *T* scores on the 12 scales used in the study. The remaining 66 were the difference between *T* scores for pairs of scales, using all combinations of the 12 scales, e.g., $L - K$, $L - F$, $L - Hs$, . . . $L - Ma$, $K - F$, $K - Hs$, . . . $K - Ma$ . . . $Sc - Ma$. This procedure was followed for both the *K* weighted and the unweighted profiles. All scores which reached the .05 level of significance with the use of the *F* ratio were retained as signs for the scale.

After the significant signs were determined, a cutting score for *schizophrenia* was arbitrarily set at the bottom of the class interval in which the mean score for schizophrenics fell. (To simplify the computations, scores were grouped in intervals of five *T* score points.) Thus a plus score would be assigned to somewhat more than 50% of the schizophrenic group on each of the significant signs and to a lesser proportion of the control group.

Results

Analysis of the *K* weighted profiles resulted in 23 difference scores and 4 single scale scores which significantly differentiated the schizophrenic from the control group. Identi-

Table 1

MMPI Scale Scores for Schizophrenic and Control Subjects (*K* Corrected)

Group	*L*	*K*	*F*	*Hs*	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Pt*	*Sc*	*Ma*
Schiz.	57.4	56.4	60.4	54.2	62.3	60.6	64.8	50.0	63.9	55.9	60.0	59.4
Control	53.8	53.5	57.6	60.9	69.8	66.5	70.2	45.8	63.0	59.7	57.4	59.0

cal analysis of the unweighted profile resulted in a slightly greater number of significant signs but yielded less satisfactory discrimination between groups, both in the validation and cross-validation samples. Consequently, further data and discussion is limited to the MMPI profile with the *K* weights applied. These mean scores are presented in Table 1.

The significant signs with the appropriate cutting score for the scoring in the schizophrenic direction are presented in Table 2.

Scoring of each profile led to the following results: The schizophrenic group had a mean score of 15.8 and the control group had a mean score of 9.3. This yielded a highly significant *F* ratio of 18.54 for 1, 87 *df*. Applying a cutting score midway between the mean scores resulted in 70% correct identification of schizophrenic *S*s and 64% correct identification of control *S*s.

Table 2

Significant Signs for Differentiating Female Schizophrenics from Controls Arranged in Categories Used for Discriminant Function Analysis[a]

	Sign	Score		Sign	Score
L	*L*	$\geqslant 53$	*Mf*	*Mf*	$\geqslant 48$
	D	$\leqslant 62$		*Hs−Mf*	$\leqslant 12$
	Hy	$\leqslant 62$		*D−Mf*	$\leqslant 12$
	Hs−L	$\leqslant 2$		*Hy−Mf*	$\leqslant 12$
	D−L	$\leqslant 7$		*Pd−Mf*	$\leqslant 17$
	Hy−L	$\leqslant 7$	*N-P*	*Pa−D*	$\geqslant -2$
	Pd−L	$\leqslant 12$		*Pt−D*	$\geqslant -2$
K	*Hs−K*	$\leqslant 2$		*Pt−Hy*	$\geqslant 3$
	Hy−K	$\leqslant 7$		*Sc−Hs*	$\geqslant 8$
	Pd−K	$\leqslant 12$		*Sc−D*	$\geqslant 3$
F	*Hs−F*	$\leqslant 2$		*Sc−Hy*	$\geqslant 3$
	D−F	$\leqslant 7$		*Sc−Pd*	$\geqslant -2$
	Hy−F	$\leqslant 2$		*Sc−Pt*	$\geqslant 3$
	Pd−F	$\leqslant 7$			

[a] For scoring individual profiles, add the number of significant signs in each category to obtain five scores. See Table 3 for appropriate weights.

Discriminant Function Analysis

Inspection of the significant signs shows that the majority fall into five logical categories: (*a*) those involving the *L* scale; (*b*) those involving the *K* scale; (*c*) those involving the *F* scale; (*d*) those involving the *Mf* scale; and (*e*) those involving the comparison of *neurotic* scales (*Hs, D, Hy, Pd*) with *psychotic* scales (*Pa, Pt, Sc*). The internal consistency of these groupings was high. Three signs do not fit into these five categories so easily. These are *D*, *Hy*, and *Sc-Pt*. These were assigned to the category with which they had the greatest correlation. *D* and *Hy* were placed in the *L* category and *Sc-Pt* was placed in the *neurotic-psychotic* (*N-P*) scales category. Table 2 has the signs arranged in these categories.

With five categories of signs available, it was feasible to analyze the scores by means of a discriminant function (Johnson, 1949). The *L* values obtained are presented in Table 3. A rough proportional value was utilized in the actual computation of results, and these weights are also presented in Table 3.

For the *K* weighted profile, the application of these weights resulted in a mean score for the schizophrenic group of 58.1 and for the control group of 32.1. Using a cutting score halfway between the means of the two groups (i.e., $\geqslant 46$ = schizophrenic identification) resulted in 68% correct identification of schizophrenic *S*s and 79% correct identification of control *S*s. This result represents a slight improvement over the unweighted sign scores.

Cross-Validation

An additional 49 *S*s were obtained at the same hospital. These were all admission cases and the diagnostic procedure coincided in time with the administration of the MMPI. Twenty-eight of these *S*s had a diagnosis of schizophrenia, and 21 had a diagnosis of neurosis or behavior disorder.

The scoring of individual profiles with the raw number of signs resulted in an over-all correct identification of *S*s of 63%. The use of the discriminant function weights applied to the same scores raised the over-all correct diagnostic identification to 79%. The *F* ratio for the weighted scores is 22.53 for 1, 47 *df*, which is significant beyond the .001 level. The greatly improved prediction leads to the conclusion that the use of discriminant function weights facilitates the prediction of schizophrenic diagnosis with these signs. Table 4 contains a summary of statistical data for all groups utilized. It will be noted that the mean and *SD*s for each group are surprisingly similar from validation to cross-validation.

MMPI data on female *S*s were solicited from a number of other NP hospitals. At the time of writing, three hospitals had responded. The results for these hospitals are combined in Table 4 for easy comparison with the other data. The *F* ratio for these scores was 41.84 for 1, 93 *df*, which is significant beyond the .001 level. Slightly reduced efficiency in discrimination is noted from the cross-validation in the original hospital.

Discussion

The schizophrenic signs obtained from a configural analysis of the MMPI profile show considerable predictive accuracy and considerable stability from sample to sample. This methodology appears to offer more promise than other approaches for predicting specific diagnoses. In addition, the methodology is easily adapted to differentiating personality variables other than diagnostic categories and easily adapted to other tests than the MMPI.

Table 3

Discriminant Function Weights for Predicting Diagnosis of Schizophrenia with Approximations Used in the Study[a]

Category of Signs	*L* Value	Approximations
L	.074	6
K	−.023	−2
F	.014	1
Mf	.012	1
NP	.092	7

[a] Multiply the approximate weight for each category in the third column above by the score in each category (from Table 2) and add for a single total score. Those scores which equal or exceed 46 were identified as schizophrenic in this study. Different cutting scores may be set in other situations should the situation demand quick discrimination of all possible patients of one or the other category.

Table 4

Discrimination Between Schizophrenic and Control Groups with Configural Analysis (Discriminant function weights—cutting score ≧46)

Groups	*N*	Mean	*SD*	% Correct Ident.
Validation				
Schizophrenic	56	58.1	21.1	68
Control	33	32.1	20.9	79
"Subtle" schizophrenic	28	52.8	19.8	64
"Obvious" schizophrenic	28	63.5	20.6	71
Cross-validation (same hospital)				
Schizophrenic	28	58.1	19.6	82
Control	21	31.3	18.9	76
Cross-validation (other hospitals)				
Schizophrenic	61	62.1	21.6	72
Control	34	33.4	18.4	76

An optimally controlled predictive study would have two features: (*a*) the MMPI results would not be known to the physician when he assigned diagnosis, and (*b*) the MMPI would have been administered at approximately the same time that diagnosis was made. Very few cases in this study fit both criteria. Insofar as the statistics are concerned, the two flaws would seem to counterbalance each other, i.e., the physician's knowledge of the test results should artificially inflate the number of correct identifications and the incongruence of time between test administration and diagnosis should artificially lower the number of correct identifications. Fortunately, two of the other hospitals used in cross-validation supplied cases in which the test was not used as part of the diagnostic process. Unfortunately, these were cases which had been on the ward for some time and probably included schizophrenics in remission and other cases whose clinical picture had changed since the diagnosis was assigned. Thus the statistic of correct identification would be artificially lowered in this instance. The over-all correct identification for this group of 48 patients was 69%. Since this is a low estimate of predictive efficiency as a result of incongruence of time between testing and diagnosis, it would appear that the physician's knowledge of test results does not grossly inflate these statistics.

The degree of statistical accuracy obtained in predicting the diagnosis of schizophrenia (approximately 75%) is adequate in itself. Intimate knowledge of the cases used makes the results more impressive. Those cases which were "misdiagnosed" seem to fall into clearly defined groups. The control *S*s identified as schizophrenic fell into several categories: (*a*) flat, unresponsive, inadequate people usually diagnosed conversion reaction or personality or character disorder; (*b*) immature and impulsive people usually diagnosed personality or character disorder; and (*c*) marginally psychotic patients who did not give sufficient evidence for a diagnosis of schizophrenia. The schizophrenic *S*s identified as nonschizophrenic fall into two general categories: (*a*) those who have had a schizophrenic psychosis but are partially or entirely recovered from it and (*b*) those who have currently active schizophrenic symptomatology but who also have many neurotic symptoms. This latter group follows a nonschizophrenic pattern of hospitalization in that they are sociable, outgoing people who make quick recoveries and are discharged as quickly as the neurotic patients. Thus in the majority of instances, there appears to be good reason for categorizing a particular patient in the wrong diagnostic group. Descriptively, the misidentified patient often resembles patients of the other category to a greater extent than he resembles the patients of his own group.

It is quickly noticed that virtually all of the significant signs involve neurotic scales of the MMPI. The psychotic scales fail to predict schizophrenic diagnoses in this sample unless their relative height is contrasted with the neurotic scales. Second most frequent in number of signs are the validation scales *L*, *K*, *F*. This appears to indicate that test-taking attitude is important in differentiating schizophrenic disorders from the neuroses and behavior disorders. It is interesting that so many predictive studies using the MMPI have ignored the validating scales except for the purpose of selecting "valid" records. It would seem that these validating scales are a source of valuable clinical information and should not be ignored.

The use of discriminant function analysis seems to have been a valuable addition to the methodology. Not only has the method increased predictive efficiency, but it permits more accurate analysis of what signs predict best. From the categories of signs, we can gain some knowledge of the aspects of test performance which differentiate the schizophrenic from the control group. The group of signs which receive the greatest weight contrasts the relative levels of neurotic and psychotic signs. Thus the female schizophrenic *S* admits to relatively more psychotic symptomatology than she does to neurotic symptomatology. On an absolute basis, she does not admit to more psychotic sympomatology than her neurotic counterpart. Second in order of importance are the *L* group of signs. It seems that the schizophrenic attempts naive deception. She tries to conceal symptomatology of a neurotic variety from herself and from others. Frequently, this goes into a denial of illness syndrome, and the patient is usually described as "lacking in insight." Third is the *F* group of signs. These contrast the neurotic scales with the *F* scale. The schizophrenic patient shows relatively more confusion and inattention as compared to the number of admitted neurotic symptoms. Once again there is not significant difference in the *F* scale alone between the groups. The fourth set of items contrasts the *Mf* scale with the neurotic scales. The schizophrenic patient shows more masculine attitudes in the absolute sense as well as relative to the amount of neurotic symptomatology expressed. The fifth group of signs contrasts score on the *K* scale, which measures a subtle form of defensiveness, with several of the neurotic scales. Taken individually, each of these signs shows that the schizophrenic female is being more defensive in relation to the number of neurotic symptoms expressed. Nevertheless, the discriminant function weight is negative, indicating that in combination with the other categories of signs, this subtle defensiveness is more predictive of nonschizophrenia. The general reason for this reversal appears to be in the fact that *L* and *K* as individual scales are highly correlated and the *L* and *K* categories of signs are also highly correlated, both measuring something akin to defensiveness or saying good things about oneself. The situation appears to be similar to what often occurs in factor analysis. After the first factor is removed, two highly correlated measures might well correlate negatively with each other. Defensiveness, as measured by the *K* scale, is believed to represent more potential for adjustment in the patient than is defensiveness as measured by the *L* scale. Perhaps a better way of expressing it is that the *L* scale seems to measure blind denial of reality, while the *K* scale appears to measure more subtle and effective defenses against reality. *K* is a more complex scale than *L* and the only unique contribution the *K* category of signs has to make when in combination with the other signs appears to be in an opposite direction from when the *K* category is taken alone. This result is in accord with clinical observation.

Summary

1. Sullivan and Welsh's approach of configural analysis was applied to the MMPI profiles of female schizophrenics and controls in an NP hospital setting. A scale of significant signs was developed with these validating groups.

2. The significant signs were grouped into five a priori, logical categories and subjected to discriminant function analysis with significantly improved predictive efficiency.

3. The weighted sign scores were then cross-validated in the same hospital and in three other hospitals with approximately 75% accuracy in discriminating schizophrenics from controls.

Received September 25, 1958.

REFERENCES

JOHNSON, P. O. *Statistical methods in research.* New York: Prentice-Hall, 1949.

MCKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The MMPI: VI. The *K* scale. *J. consult. Psychol.,* 1948, **12**, 20–31.

SULLIVAN, P. L., & WELSH, G. S. A technique for objective configural analysis of MMPI profiles. *J. consult. Psychol.,* 1952, **16**, 383–388.

TAULBEE, E. S., & SISSON, B. D. Configurational analysis of MMPI profiles of psychiatric groups. *J. consult. Psychol.,* 1957, **21**, 413–417.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

ANXIETY INDICES IN SAME–SEX DRAWINGS OF PSYCHIATRIC PATIENTS WITH HIGH AND LOW MAS SCORES[1]

THOMAS E. HOYT[2] AND MARTIN R. BARON

Kent State University

Since the publication of Buck's (1948, 1950) and Machover's (1949) monographs, the House-Tree-Person (H-T-P) and Draw-a-Person (DAP) have become instruments in common use by the clinical psychologist. This use has been accompanied by many studies concerning the reliability and validity of the projective devices. A recent article (Swenson, 1957) ably summarizes the evidence to date, most of which leaves serious doubts regarding the validity of many hypotheses concerning the meaning of drawing characteristics. Despite these findings, many writers feel that the technique holds promise.

In suggesting possible uses of the H-T-P, Buck (1948) has pointed out some characteristics of the "person" drawing which are often noted among patients clinically diagnosed as anxious. Similarly, Machover (1949) has specified in the use of the DAP those elements in pictures which in her clinical experience are indicative of anxiety. In addition, they both mention characteristics of drawings which they believe indicate personality characteristics generally thought to be associated with anxiety. The purpose of this study is to compare the frequencies of anxiety indices in psychiatric patients classified as anxious or nonanxious on Taylor's Manifest Anxiety Scale (MAS) (Taylor, 1953).

Indices Investigated

Table 1 lists all the characteristics of human figure drawings which Buck and/or Machover believe show anxiety; Table 2 lists characteristics indicated by Buck and/or Machover to be related to personality and considered by the principal author to be worth checking as potential anxiety indicators. All characteristics listed in Tables 1 and 2 are investigated in the present study.

Evidence from other investigations regarding the validity of these indices will be considered at a later point.

Criterion Measure

The objective measure of anxiety with which the various anxiety indicators listed in Tables 1 and 2 are compared in this study is the Manifest Anxiety scale (Taylor, 1953). Designed not as a clinical tool but to select subjects for experimental groups, the test has been employed to select extreme groups presumably in regard to anxiety; performance measures for these groups are usually found to differ (Taylor, 1953). In accordance with the use typically made of the MAS, high and low anxiety groups were selected in the present study, and a tally then made of the

Table 1

Characteristics of "Person" Drawings Investigated in the Present Study and Considered by Buck and/or Machover to be Indicators of Anxiety

Characteristics of Drawing	Buck Reference	Machover Reference
Placement (upper left-hand corner)	(1948, p. 372)	
Type of line (faint)	(1948, p. 379)	
Reinforcement	(1948, p. 368)	(1949, p. 110)
Shading	(1948, p. 377)	(1949, p. 98)
Erasing		(1949, pp. 98, 110)

---

[1] This study constitutes a portion of a thesis performed under the direction of the second author in partial fulfillment of requirements for the M.A. degree at Kent State University, August, 1955.

[2] Formerly at State Bureau of Vocational Rehabilitation, Youngstown, Ohio. Now deceased.

Table 2

Characteristics of "Person" Drawings Investigated in the Present Study and Considered by Buck and/or Machover to be Indicators of Personality Attributes Other Than Anxiety

Characteristic of Drawing	Personality Attribute	Buck Reference	Machover Reference
[a]Type of line (faint)	Timid; Uncertain; Self-effacing		(1949, pp. 95–96)
[a]Erasing	Conflict	(1948, p. 380)	
Size (small)	Inadequacy	(1948, p. 370) (1950, p. 22)	
Relative size of head to figure	Relative emphasis on Intellectual Power		(1949, p. 36)
Size of head	Social Dominance and Control of Body Impulses		(1949, p. 36)
Omissions	Helplessness	(1948, p. 367)	
	Guilt		(1949, p. 62)
	Conflict		(1949, p. 97)
Body area out of proportion	Shallow Emotionality		(1949, p. 91)
	Lack of Insight		(1949, p. 91)
	Poor Reasoning		(1949, p. 91)

[a] The characteristics *type of line* and *erasing* are also included in Table 1.

number of anxious and nonanxious subjects displaying the various anxiety indices in their drawings.

Method

Subjects

Same-sex drawings of institutionalized women are used as the basis of this study. The *S*s were neurotic and psychotic women who were new admissions to Woodside Receiving Hospital at Youngstown, Ohio between January and June 1955. Tests were administered almost weekly during this period in order that the *S*s would be tested shortly after admission. The selective factors required that they be: (*a*) between 20 and 65 years of age; (*b*) able to read and write; and (*c*) in a sufficient state of awareness and stability (as judged by the resident psychologist) to participate as *S*s.

A total of 125 women was tested, but for various reasons, such as illness, incomplete drawings, etc., 13 were eliminated. The 112 *S*s were divided into three groups according to their MAS scores. The 27% with the highest scores and the 27% with the lowest scores were used for this study. This resulted in 60 *S*s, one group of 30 with a score on the MAS of 1 to 12, to be called "low anxiety," and one group of 30 with a score on the MAS of 25 to 43, to be called "high anxiety."

Procedure

Testing was done in a group situation, the number of *S*s in the group being from 5 to 15 women. The MAS was administered to the *S*s, then the *same-sex drawing* test. Following this, a group Rorschach was given. Finally, *S*s were asked to draw a second figure of their own sex. The data herein considered concern only the first drawings.

Scoring and Reliability

An attempt reported elsewhere (Hoyt, 1955) was made to devise a refined scoring manual to measure the various possible anxiety indices from the drawings. All drawings were first coded so that the examiner would not know which pictures were drawn by high and which by low anxiety *S*s, and then scored on one index at a time. Score values usually from zero to three were assigned.[3] Reliability of this scoring procedure

---

[3] The aim of Hoyt's (1955) original study was to determine the existence of significant relationships between graduated amounts of various anxiety indices and MAS scores. To do this, he utilized a pilot group to determine for each index which way of scoring amounts of "presence" would correlate

Table 3

Proportion of Anxious and Nonanxious *S*s Displaying Various Characteristics in Their Drawings, with Chi Square Values and Levels of Confidence

Characteristic of Drawing	Proportion Displaying Characteristic: Anxious	Proportion Displaying Characteristic: Non-anxious	Chi Square	Level of Confidence
Placement (upper left-hand corner)	.433	.167	3.88[a]	.05
Type of line (faint)	.167	.133	0.00[a]	1.00
Reinforcement	.700	.533	1.20	.30
Shading	.433	.400	.07	.80
Erasing	.833	.867	0.00[a]	1.00
Size	.433	.167	3.88[a]	.05
Relative size of head to figure	.533	.633	.62	.50
Size of head	.433	.167	.94[a]	.40
Omissions	.400	.233	1.22[a]	.30
Body area out of proportion	.467	.533	.26	.70

[a] Corrected for continuity.

was determined by computing for each index the mean absolute difference between the scores assigned by the principal investigator and those independently assigned by a second psychologist for 112 *S*s. Since the mean differences were generally small (only two were larger than .111), the reliability of this scoring procedure was good. For the purposes of the present report, all anxiety indices have been rescored into dichotomies such as presence or absence. Since most instances of disagreement between the two scorers occurred regarding amounts of presence, utilizing two categories serves to increase reliability. Following the rescoring on each index, the data were analyzed by finding the significance of difference between the high anxiety group and the low anxiety group on each index by the use of the chi-square test. Yate's correction was employed on all tables where there was an expected value less than 10.

Results

The proportion of anxious and nonanxious *S*s displaying the various indices in their drawings, the chi-square values for each index and the levels of confidence appear in Table 3. In view of the fact that virtually all *S*s shaded the hair of the person in the drawings, *shading* was counted only for body shading. Two of the indices were found to differentiate significantly between high and low anxiety groups on the basis of the MAS criterion.

Placement in upper left-hand corner and small *size* were found to be the most valid indices according to the MAS criterion with chi-square values significant at the .05 level of confidence. *Reinforcement* and *omissions* were found to differentiate to some degree between high and low anxiety groups, but the chi-square values would be exceeded 30% of the time by chance alone. The other indices were found to discriminate not at all between high and low MAS groups. Thus, of 10 anxiety indices, only two are found to be valid at the .05 level when separately employed utilizing the MAS as a criterion of anxiety.

It is perhaps worthwhile at this point to re-examine Buck's and Machover's hypotheses. Are they saying that the presence of *shading* (as an example) is indicative of anxiety? If this is so, then our chi-square test of each index separately has been appropriate. Or are they saying that anxiety may be indicated in a number of ways; that the presence of any one of many indices predicts anxiety? If the latter is true, then the results reported in Table 3 are not a fair test of their hypotheses. The data reported in Tables 4 and 5 are relevant to the hypothesis that that any *S*, whose drawing has any one of the characteristics indicated by Buck and Machover to be indices of anxiety, is more likely to be anxious than is an *S* whose drawing lacks any of these characteristics.

Machover's (1949) indices are *reinforcement*, *shading*, and *erasing*. None of these, it will be recalled, is significantly related to the criterion. A count of the number of high and low anxiety *S*s, shown in Table 4, whose pictures contained any one of these indices, showed 87% of the anxious and 93% of the

---

best with the MAS. He hoped to test the hypothesized indices and scoring procedures on a new independent sample; however, a new sample large enough to perform an adequate test of the utility of his scoring procedure could not be obtained. Because of this unfortunate characteristic of Hoyt's (1955) design, his scores have been dichotomized into what might well be considered a priori categories.

nonanxious *S*s to have displayed one of these characteristics in their drawings. Because the proportion of cases showing at least one of these characteristics is so close to 1.00, chi-square computation is unjustified. It is clear from the figures, however, that the presence of one of Machover's indices is not found more frequently in anxious than nonanxious *S*s.

The anxiety indices mentioned by Buck (1948, 1950) are *placement* in upper left-hand corner, light *type of line*, *reinforcement*, and *shading*. Of these, the first is significantly related to our MAS criterion. A contingency table (Table 5) was constructed to determine if the presence of any of Buck's four indices was more likely to be found in drawing of high MAS than those of low MAS *S*s. It is notable here, too, that a very high proportion of both anxious and nonanxious *S*s displayed at least one of Buck's indices in their drawings. Clearly, the presence of any one of his indices does not differentiate high and low MAS groups; chi square equals zero.

Discussion

It was apparent to the authors at the outset of this study that there might very well be little correspondence between anxiety, clinically diagnosed, and manifest anxiety, as measured by Taylor's MAS scale. Some might argue that skilled interpretations of drawings of the human figure would be a better criterion of anxiety than the MAS, and thus a correlation between MAS and drawing interpretations would indicate the validity of the MAS. Considering that the present data are simply correlational, it is noteworthy that 8 of the 10 anxiety indices investigated in this study were found not significantly related to the MAS criterion. It is perhaps more worthy of note, on the other hand, that two of the indices, one of them specifically indicated by Buck as an anxiety indicator, *are* significantly related to the criterion. Since it is unlikely that 2 of the 10 indices could be related to the criterion (at the .05 level of confidence) by chance, the authors take the position that the interpretations of the drawings and MAS indicate more in common than might be expected on a chance basis.

Table 4

Proportion of Anxious and Nonanxious *S*s Displaying and Not Displaying Any One of Machover's Anxiety Indices (Reinforcement, Shading, and Erasing) in Their Drawings

	Displaying	Not Displaying
Anxious	.867	.133
Nonanxious	.933	.067

Table 5

Proportion of Anxious and Nonanxious *S*s Displaying and Not Displaying Any One of Buck's Anxiety Indices (Placement, Type of Line, Reinforcement, and Shading)

	Displaying	Not Displaying
Anxious	.867	.133
Nonanxious	.833	.167

Note.—($x^2 = 0$).

It is perhaps worthwhile at this time to compare the evidence obtained in this study with other findings described in the literature. Table 6 is presented to facilitate the comparison.

Here it can be seen at a glance that, according to evidence prior to this study, only two characteristics are related to anxiety: *size* and *size of head*. Of these, size was also found significantly related to our MAS criterion. In addition, placement was found before not to be related to anxiety, whereas in the present study it *was* related to MAS scores. Thus, the findings of the present study are partially in agreement with those reported in the literature, and point up the relevance of placement where such relevance had not been previously confirmed.

Using the MAS criterion, it becomes apparent, as shown in Tables 4 and 5, that the presence of any one of the anxiety indices (either of Machover or Buck) does indicate anxiety; however, it is also true that they indicate nonanxiety just as often. One possible interpretation of this finding is that in clinical experience one may be tempted to remember only positive evidence and as a consequence to continue employing an instrument.

Table 6

Characteristics of "Person" Drawings Investigated in This Study as Possible Anxiety Indicators with the Nature of Prior Findings and Present Findings Indicated Separately

Characteristics	Prior Findings	Present Findings
Placement	negative (Goodman & Kotkov, 1953)	positive
Type of line	negative (Royal, 1949)	negative
Reinforcement	negative (Royal, 1949)	negative
Shading	negative (Royal, 1949)	negative
Erasing	negative (Goldworth, 1950; Royal, 1949)	negative
Size	positive (Lehner & Gunderson, 1953)	positive
Size of head and relative size of head to figure	positive (Goldworth, 1950)	negative
Omissions	positive (Goldworth, 1950) negative (Goldworth, 1950; Royal, 1949)	negative
Body areas out of proportion	positive (Goldworth, 1950) negative (Prater, 1950; Royal, 1949)	negative

In the light of the present evidence, it is suggested that attempts be made to determine if the presence of any of these indices is *more* common among *S*s clinically diagnosed as anxious than among those diagnosed as nonanxious. If data comparable to those in Tables 4 and 5 are obtained, using clinical criteria, virtually all psychiatric patients are anxious, and the presence of these indices is not particularly worthy of note except as possible sources of suggestion to the therapist regarding possible anxiety sources.

Summary

The purpose of this study was to determine whether various indices clinically employed by Buck and/or Machover to diagnose anxiety were valid, using the Taylor Manifest Anxiety Scale (MAS) as a criterion. Other indices were also investigated.

The indices investigated were *placement* in the upper left-hand corner of the page, faint *type of line*, *reinforcement* (i.e., retracing), *shading*, *erasing*, *size*, *relative size of head to figure*, *size of head*, *omissions*, and body areas out of *proportion*.

The 112 *S*s who were employed in this study were female admissions to a mental receiving hospital. These *S*s were given the MAS and then asked to draw a figure of their own sex. The highest and lowest 27% of these 112 *S*s, as determined by the MAS, were classed respectively as anxious and nonanxious and the drawings of these 60 *S*s independently scored for presence or absence of each index.

Findings showed *placement* and *size* to be significantly related to MAS scores. None of the other characteristics significantly differentiated between the anxious and nonanxious patients.

Received October 2, 1958.

REFERENCES

Buck, J. N. The H-T-P Technique: A qualitative and quantitative scoring manual. *J. clin. Psychol.*, 1948, **4**, 317–396.

Buck, J. N. Administration and interpretation of the H-T-P Test. *Proc. H-T-P Workshop, U. S. Vet. Adm. Hosp.*, Richmond, Va., 1950.

Goldworth, S. A comparative study of the drawings of a man and a woman done by normal, neurotic, schizophrenic and brain-damaged individuals. Unpublished doctoral dissertation, Univer. of Pittsburgh, 1950.

Goodman, M., & Kotkov, B. Predictions of trait ranks from Draw-A-Person measurements of obese and non-obese women. *J. clin. Psychol.*, 1953, **9**, 365–367.

Hoyt, T. F. Same-sex drawings as a measure of anxiety in institutionalized women. Unpublished master's thesis, Kent State Univer., 1955.

Lehner, G. F. J., & Gunderson, E. K. Height relationships on the Draw-A-Person test. *J. Pers.*, 1953, **21**, 392–399.

Machover, Karen. *Personality projection in the drawing of the human figure.* Springfield, Ill.: Charles C Thomas, 1949.

Prater, G. F. A comparison of the head and body size in the drawing of the human figure by hemiplegic and nonhemiplegic persons. Unpublished master's thesis, Univer. of Kentucky, 1950.

Royal, R. E. Drawing characteristics of neurotic patients using a drawing-of-a-man-and-woman technique. *J. clin. Psychol.*, 1949, **5**, 392–395.

Swenson, C. H., Jr. Empirical evaluations of human figure drawings. *Psychol. Bull.*, 1957, **54**, 431–466.

Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

NOMOGRAPHS FOR COMPUTING THE "VALIDITY" OF WISC OR WECHSLER–BELLEVUE SHORT FORMS

CECIL BRIDGES

Department of Pupil Services, Oklahoma City Public Schools

In the course of the selection of a short form of the Wechsler Intelligence Scale for Children (WISC) to be used as a screening instrument in special education, the problem of calculating numerous correlations of possible short form subtest combinations with total score arose. McNemar (1950) has shown that the correlation between abbreviated forms of the Wechsler-Bellevue (WB) Scale and total score may be calculated from correlation matrices presented by Wechsler (1944) in his manual. This is also true of the WISC (Wechsler, 1949).

There are several instances in which computation of such "validity" coefficients may be useful. For instance, in selecting a short form screening instrument from subtests, factors other than WB short form correlation with total score might enter into the selection (ease of administration, etc.), and the "validity" of the chosen battery might not be among those given by McNemar. In the case of the WISC no such coefficients are available.

In using McNemar's formula, the calculations require a desk calculator if reasonable speed is desired. The nomographs presented here were devised to eliminate all calculations except addition necessary to obtain the sum of the correlations of the tests and the sum of their intercorrelations. One nomograph is used for computing the short form correlation with total score when the short form consists of two, three, or four subtests. A second nomograph is used for computing the correlation when five, six, or seven subtests are used. With the use of a nomograph, calculation is considerably faster than with a desk calculator.

To use the nomograph, first find the sum of the *r*'s for each selected subtest from the correlation matrix presented in Wechsler's manual (there will be nine of these *r*'s for each test) and add these sums.[1] Next find the sum of the intercorrelations between the selected subtests.

The horizontal scales on the nomograph correspond to the sum of the *r*'s (the first quantity mentioned above). The vertical scales on the nomographs correspond to the sum of the intercorrelations of the chosen tests. The curved lines in the body of the nomograph correspond to the correlation between the short form and total score. Thus, to find the correlation of the short form with total score, find a point on the appropriate horizontal scale corresponding to the sum of the *r*'s (interpolate if necessary) and lay a straight edge through this point perpendicular to the scale. Find a point on the appropriate vertical scale corresponding to the sum of the intercorrelations, and lay a straight edge through this point perpendicular to the scale. If the intersection of the two perpendiculars

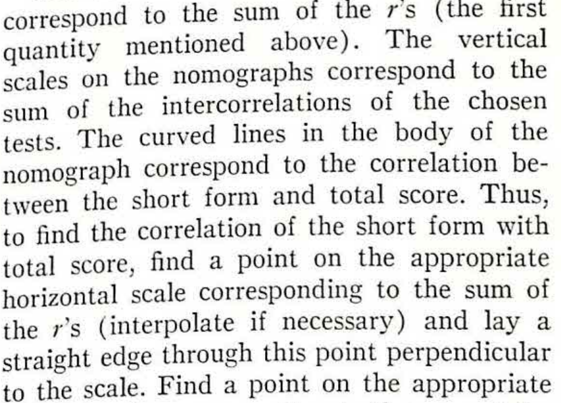


Fig. 1. Nomograph for computing the correlation of short form with total score when two to four subtests are used.

---

[1] Exclude WISC *mazes* and *digit span*.

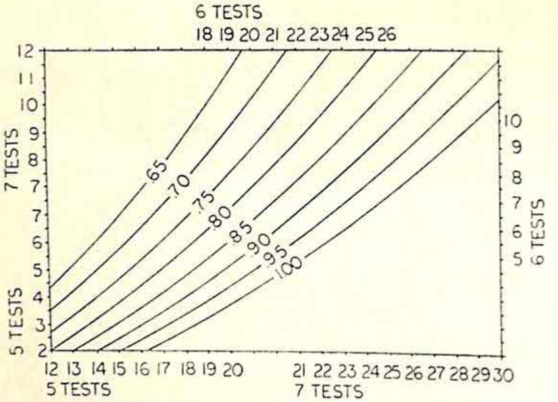


Fig. 2. Nomograph for computing the correlation of short form with total score when five to seven subtests are used.

falls on one of the contours of equal correlation, the value of r with which that line is labelled is the correlation of the short form with the total. If the point of intersection falls between contours of equal correlation, interpolation is necessary. The correlation found will usually be correct in the second decimal place if interpolation is done carefully.[2] The nomographs are strictly applicable only to the populations of Wechsler's (1944) Table 41 and Wechsler's (1949) Table V. Because the subtest correlations are not perfectly reliable, there is some capitalization on chance when a battery of subtests is selected for maximum correlation with the total, and a slight shrinkage of this correlation would be expected in cross-validation.

Because of a small arithmetical error in McNemar's special formula (Howard, 1958), WB "validities" obtained with these nomographs will be slightly higher than corresponding "validities" presented in McNemar's original article.

Received October 7, 1958.

REFERENCES

Howard, W. A note on McNemar's "On abbreviated Wechsler-Bellevue scales." *J. consult. Psychol.*, 1958, **22**, 414.

McNemar, Q. On abbreviated Wechsler-Bellevue scales. *J. consult. Psychol.*, 1950, **14**, 79–81.

Wechsler, D. *The measurement of adult intelligence.* (3rd ed.) Baltimore: Williams & Wilkins, 1944.

Wechsler, D. *Wechsler intelligence scale for children: Manual.* New York: Psychological Corp., 1949.

---

[2] There is a negligibly small constant error of about ¼ of 1% involved in the use of a single set of nomographs for both tests. The use of a single set of nomographs is possible because the quantity in the first radical of the denominator of McNemar's equation is approximately the same for the WISC Table V (Wechsler, 1949) and WB Table 41 (Wechsler, 1944).

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

FORCED ASSOCIATIONS, SYMBOLISM, AND RORSCHACH CONSTRUCTS

JOSEPH F. RYCHLAK

Washington State University

An interesting and problematical aspect of Rorschach test interpretation is that of content analysis. Is it legitimate to base hypotheses about personality on the content of a subject's perception, or should the clinician limit himself to the more rigorous procedure of formal scoring? Piotrowski (1957, p. 324) notes that the latter practice is on the decline and that psychologists are relying increasingly on content analysis.

Rorschach (1943, p. 207) suggested that content of inkblot interpretation might have a meaning of its own, but believed this to be determined primarily by significant relationships existing between form and content. A review of representative texts on current Rorschach interpretation (Ames, Learned, Metraux, & Walker, 1954; Beck, 1952; Freud, 1938; Lindner, 1952; Rorschach, 1943; Sarason, 1954; Schafer, 1954) reveals that none actually denies the possibility of content responses carrying a meaning of their own. Usually the author cautions against hasty symbolic interpretations and strongly urges that content is only one of several dimensions—and probably a lesser one—to consider in any given protocol. Often the context of the content response is stressed, with the implication that the meaning of the latter varies with the former. Following such admonitions, however, several authors go on to hypothesize a given meaning for a given content response (Allen, 1954, p. 85; Halpern, 1953, p. 37; Klopfer, Ainsworth, Klopfer, & Holt, 1954, p. 385; Phillips & Smith, 1953, p. 120). Such proposals may be related to a specific area of a card, but this is not always the case.

Where do these interpretations come from? Clinical experience is usually cited as the major source. However, since clinicians do not work in theoretical vacuums it is not surprising to find that many of the symbolic interpretations made on the basis of content stem from the psychoanalytical literature (Piotrowski, 1957, p. 324). A more basic question is how valid is it to generalize a content interpretation based upon past clinical cases to the present one? Do people really ever see the same blot or, more profoundly, the same content? How can one check on the meaning of a content response? We are probably a long way from any conclusive method for the testing of content meaning. Osgood (1952) has presented a method which may eventually develop into an adequate test for the meaning of symbols. Some would argue that this will never be possible on a large scale and that clinical evaluation will remain the best bet. Although agreeing in spirit with the latter position, the author would like to describe a method for testing, in gross fashion, the relationship between a given construct and a given meaning.[1]

The method employs a simple forced association between arbitrarily chosen Rorschach constructs (Clouds, Fire, etc.) and a series of six "meanings" (Love, Anger, etc.) often assigned to these constructs (rightly or wrongly) by clinicians. No claim is made as to the actual meaning of a construct; all the method aims to do is assess whether or not people can agree in their choices between some of the meanings *often assigned* by clinicians. It is also of interest to see, given a choice between various meanings, which one will be most frequently chosen.

---

[1] The author would like to thank the following people for their help in arranging *S*s and analyzing the data: Julian C. Davis, Florida State Hospital; Barron B. Scarborough and Charles Moos of Florida State University; and John W. Neff, Clinical Director, Pueblo, Colorado Mental Hygiene Center.

A construct is seen as a cognitive abstraction from sensory stimulation, which may have more than one circumscribed pattern of stimulation as a referent. The construct "foggy," for example, can refer to an atmospheric condition or a state of mind. Since the former association is more frequent, we use the term "symbolical" to describe the situation when the latter association is utilized. Osgood (1952, p. 200) differentiates *signs* from *objects,* and the distinction can help in clarification here. An object is any pattern of stimulation (from a hammer to a gust of wind in the face) which evokes reactions on the part of an organism. A sign is any pattern of stimulation *other than* the object which evokes reactions relevant to the object. The theoretical position taken here is that Rorschach constructs are such signs. If Rocks symbolize Security (Halpern, 1953, p. 37) what the clinician must mean in this suggestion is that a meaning is conveyed by the construct Rock. He—the clinician—has learned to identify this less frequent association which people have between feelings of security (object) and rocks (sign). The process of projection would be the opposite side of the coin: a person has need of security (object), hence presumably chooses from the amorphous stimulus of the inkblot a symbol (sign) typifying the need.

The present study is based upon a social psychological rationale. It is assumed that if constructs chosen from the Rorschach have the meanings often ascribed to them by clinicians, then a polling of assorted groups should reflect these meanings. This is so because cultural influences must carry the consistency from person to person and, except for minor subcultural differences, broad similarities between groups should prevail. As a theory of symbolism, the point of view here espoused comes closer to that of Hall (1953) than to the more traditional psychoanalytical view, although Freud certainly recognized the cognitive as well as defense functions of the symbol.

Method

Instrument

A forced association test, consisting of three pages and devised for group use, was administered to the *S*s. Each page dealt with the following 12 constructs: Boots, Smoke, Bear (animal), Mask, Fur, Fire, Clouds, Rocks, Hair, Bat (animal), Island, Mountains. A description of each page follows:

Page 1: Constructs were ranked at the left-hand side of the page and to their right were two columns of blank spaces. The columns of blanks were titled Positive Meaning and Negative Meaning, making it possible for an *S* to assign a positive or negative valence to each construct.

Page 2: Constructs were arranged so that each was followed by three meanings: Ambition, Love, and Security. A check space preceded each meaning so that a construct could be associated with one of the three alternatives by simply placing a check before the meaning "coming closest" to the construct.

Page 3: Similar to page 2, except that each construct was followed by the meanings of Depression, Fear, and Anger.

The Rorschach constructs were selected at random from those frequently reported by *S*s taking the test. Two points need to be made concerning the list of meanings. First of all, as discussed in the introduction, in another context they could serve as constructs. They are arbitrarily "meanings" here only because of their representativeness as frequently made interpretations of the constructs. For this reason, the *E* defined each meaning in the experimental instructions to *S*s, but did not define the constructs. Secondly, the meanings were chosen as a result of extensive pretesting during which it was discovered that *S*s found it difficult to differentiate between meanings having similar connotations, such as Anger and Hatred, or Sex and Love. Of course, it is possible to think of Love as an Ambition, or reflecting Security, but it was found through interviewing *S*s that the meanings eventually chosen were less likely to be confused than those of, e.g., Anxiety and Fear. Psychologists may have a clear conception of the differences between the latter, but *S*s rarely do.

Consequently, the final instructions were so phrased as to make a broad inclusion of connotations under a meaning possible. Page 2 instructions included the following brief overview by the *E:*

> Ambition meaning of course the desire to get ahead or improve one's circumstances . . . . Love refers to all kinds of warm emotions or feelings toward other people or other things . . . and Security meaning the feeling of safety and protection one experiences in certain situations.

Page 3 had the following clarifying remarks:

> Depression meaning of course a feeling of sadness and gloom, a sort of "what's the use?" notion . . . . Fear means the feeling of fright we get when something scares us or worries us and makes us anxious . . . and Anger is the kind of irritable feeling of being mad or burned-up at someone or something.

No suggestions regarding the definition of constructs were given, except to clarify that two of them (Bear and Bat) referred to animals.

Thirty-five *S*s (20 female, 15 male) attending an undergraduate psychology course were given a one-week test-retest reliability check, evaluated with the binomial expansion (Edwards, 1957, pp. 43–48). The mean CA of these *S*s was 22.6, and their mean educational level was 14.9. In order to reach the .05 level of significance, an *S* must match his initial choice, on retest, at least nine times for page 1, and at least six times for pages 2 and 3. On this basis, 97% of all *S*s reached the .05 level on page 1, 97% reached it on page 2, and 86% reached it on page 3. To demonstrate the upper limits of reliability, a more rigorous level of significance can be pointed to. For example, 51% of the *S*s had 11 or more matches (.003) on page 1; 74% had eight or more (.007) on page 2; and 40% had eight or more (.007) on page 3. Thus, it appears that constructs and meanings can be associated with an acceptable degree of consistency.

Subjects

The following groups of *S*s were tested ($N = 160$):

University Students: $n = 67$ (30 female, 37 male); mean CA 21.4; mean education 14.2

Male Extension Students: $n = 22$; mean CA 27.3; mean education 12.8

Female Extension Students: $n = 25$; mean CA 40.2; mean education 14.8

Male Mental Patients (MPs): $n = 23$; mean CA 32.0; mean education 11.6

Female MPs: $n = 23$; mean CA 40.1; mean education 11.8

The university students were drawn from introductory psychology classes. Male extension students were USAF personnel, attending night classes at a nearby airbase. The female extension students included several preschool teachers. The MPs were in residence on the receiving wards of a State Hospital. Although some were bizarre, in general they were not the severely regressed type of patient. The only criteria of selection adhered to in choosing the MPs were that they have either (*a*) a high school education or (*b*) two years of high school and score within the average range of intelligence (or above) on a standard IQ test (Wechsler or Binet). All forms of diagnostic categories were included, from character disorders through neuroses and psychoses.

Procedure

The study was structured for *S*s as a "survey of concept meaning." They were told that the *E* was interested in knowing what most people "associate—or connect—with certain concepts." All testing was done in groups, but the MP sample involved much more intimate groupings of two to eight *S*s. After checking a positive or negative valence on page 1—presented through an example in the instructions as roughly equivalent to liking or disliking the construct—*S*s went on to complete pages 2 and 3 as a group (i.e., all *S*s completed each page before another was begun).

Only one of the three alternative meanings could be checked for each construct. *S*s were told that there were no right or wrong answers, and that if none of the three meanings fitted exactly they were to check the one coming closest to what the construct suggested. Questions and complaints were occasionally raised, but on the whole, test administration ran smoothly. Only two of the original group of MPs had to be dropped from the sample because of a refusal to cooperate.

Results

The first question to consider is—are there consistencies in forced associations between constructs and meanings? Table 1 contains the valances and meanings most frequently associated to the constructs by the total sample, evaluated with chi square (Edwards, 1957). The percentage of all *S*s ($N = 160$) choosing a valence and meaning is listed. Note that in some cases, although statistical significance exists (.05 level or greater), no clear-cut choice between meanings associated to a given construct is possible.

Table 1

Percentage of Choices: Valences and Primary Associations ($N = 160$)

	Valence		Page 2 Primary Association		Page 3 Primary Association	
Construct	Value	% of N Choosing	Mng.	% of N Choosing	Mng.	% of N Choosing
Boots	Pos.	74**	Sec.	61**	Anger Depr.	36 40 }*
Smoke	N.S.		N.S.		Fear	63**
Bear	Neg.	58*	N.S.		Fear	71**
Mask	Neg.	70**	Amb. Sec.	42 41 }**	Fear Depr.	44 36 }**
Fur	Pos.	86**	N.S.		Depr.	52**
Fire	N.S.		Amb. Sec.	38 41 }**	Fear	78**
Clouds	Pos.	83**	Love	49**	Depr.	77**
Rocks	Pos.	70**	Sec.	63**	N.S.	
Hair	Pos.	82**	Love	64**	Depr.	53**
Bat	Neg.	82**	Amb.	57**	Fear	69**
Island	Pos.	86**	Love	50**	Depr.	63**
Mountains	Pos.	91**	N.S.		Fear Depr.	39 43 }**

Note.—N.S. means "not significant."
* Significant between .05 and .01 levels.
** Significant at or beyond .01 level.

A reasonable second question revolves about the influence of valences upon associations. Did *S*s viewing a construct as either positive (i.e., roughly equivalent with "liking") or negative significantly alter the pattern of associations noted when this distinction was not taken into consideration? Table 2 contains a valence breakdown, giving the number of *S*s (n) who assigned either positive or negative to a construct and, of these, the percentage who selected the listed meaning in their forced associations. Note that the meanings most frequently associated to a construct are remarkably similar in both tables. Moreover, the percentage of *S*s choosing a given meaning remains roughly equivalent. Interesting differences can be pointed to. For example, the construct Bat no longer retains its association with Ambition when we consider the choices of only those *S*s (28) who checked Bat as having a positive valence. A finding of this sort suggests that, at least in some cases, significant differences resulted because *S*s were choosing the alternative meaning *least different* than a construct, and did this with consistency.

The next point to consider concerns any group differences that might be expected on the basis of group classification. Do MPs have "private meanings" differing from the other groups which might account for some of the variance noted in the tables? It was decided that only the broad distinctions of MP-normal and male-female would be of immediate significance in the present investigation. Finer breakdowns would not be necessary unless significantly different trends were noted in the larger comparisons. Consequently, an impulse to contrast each group was purposely checked.

Comparisons of the specified subsamples revealed no major differences with regard to predominant valence or with regard to the most heavily favored associations. Only 12 of the 72 chi squares computed reached significance (.05 level or greater), and of these only three displayed a difference in predominant valence or primary association (i.e., meaning chosen most frequently by either of the groups under comparison).

For example, a typical finding (.01 level) was that although MPs (63%) and normals (62%) both associated Security to Rocks, normals (32%) were more likely to also associate Ambition, whereas MPs (22%) tended to choose Love as a secondary association to Rocks. An example of a relative valence difference (.01 level) is that women (94%) are more likely to assign a positive valence to Fur than men (80%). Such findings, although certainly suggestive, cannot be emphasized because of the gross nature of the polling.

The three comparisons having differences in valence choice or primary association are as follows: MPs (76%) are more likely to attach a negative valence to Fire than normals (48%) (.01 level). MPs (52%) tend to associate Fear to Mountain most frequently, whereas normals (43%) associate Depression most frequently (.02 level). Women (62%) associate Love to Clouds, whereas men (49%) are more likely to associate Ambition (.01 level).

A final consideration of interest in a survey of this sort concerns the frequency of meaning categories chosen, disregarding the particular construct to which they were to be associated. Do the experimental groups make equal use of all meanings, or do some favor a particular category? Additional tests of significance again revealed a fair degree of consistency concerning primary associations, yet some differences were noted. Male MPs chose Security (42%) and Love (34%) most frequently, whereas normal men selected Ambition (38%) and Security (37%) most frequently (.01 level). No comparable differences were found when contrasting female MPs and normal females. However, sex differences did exist. Women, as a group, chose Anger (18%) and Fear (47%) as extremes, whereas the men had less of a contrast between extremes with Anger (24%) and Depression (39%) as their low and high choices (.01 level).

Discussion

Leaving aside the issue of which particular meaning was associated with which particular construct, it seems fairly apparent that the major prediction of this investigation has received some verification. A consistency in forced association was noted across such lines as sex and mental health. Only three actual reversals in valence or primary association were noted. The slight inconsistencies between groups probably can be attributed to chance, although they are explicable when taking the groups into consideration. That is, women assigning a positive valence to Fur more frequently than men is understandable, considering the female status implications of

Table 2

Percentage of Choices: Valences Breakdown ($N = 160$)

	Positive Valence (n varies)			Negative Valence (n varies)		
Construct	n	Mng.	% of n Choosing	n	Mng.	% of n Choosing
Boots	118	Sec.	62**	42	Depr.	53*
Smoke	74	N.S.		86	Fear	65**
Bear	67	N.S.		93	Fear	77**
Mask	48	Amb. Sec.	35 } ** 47 }	112	Fear Depr.	41 } ** 38 }
Fur	138	N.S.		22	Depr.	64**
Fire	70	Sec.	49*	90	Fear	82**
Clouds	133	Love	51**	27	Depr.	88**
Rocks	112	Sec.	61**	48	N.S.	
Hair	132	Love	65**	28	Depr.	61**
Bat	28	N.S.		132	Fear	71**
Island	138	Love	51**	22	Depr.	64**
Mountains	146	N.S.		14	N.S.	

Note.—N.S. means "not significant."
* Significant between .05 and .01 levels.
** Significant at or beyond .01 level.

a fur coat. Or, the fact that male MPs use Love more frequently than normal males (who use Ambition) when associating to the constructs makes sense in light of the former's circumstances as hospitalized patients in need of care, etc.

On a more theoretical plane, the results are in line with a sociopsychological approach. If constructs have a symbolic function then, unless the clinician has truly extensive time in which to interview an individual to determine unique aspects of his symbolic activity, it must be assumed that this is a socially learned function, carried by cultural influences. The test or dream interpreter, as a member of this culture and by assuming the proper set to observe, can identify these "symbols" in the fashion outlined in the introduction of this paper. By studying the tables the reader can assess for himself the accuracy with which his particular bias of symbolic interpretation meets the findings. Suffice to say that many common interpretations hold. This study can say nothing about the *actual* meaning of a construct.

Summary

The present research approaches the problem of construct symbolism with a sociopsychological rationale. It is assumed that, if constructs chosen from the Rorschach have the content meanings often ascribed to them by clinicians, then a polling of assorted groups should reflect such consistencies in forced associations. Assorted groups of *S*s polled are found to reflect certain consistencies in their forced associations, cutting across such lines as sex and mental health. The major prediction is considered verified.

Received October 13, 1958.

REFERENCES

ALLEN, R. M. *Elements of Rorschach interpretation.* New York: International Universities Press, 1954.

AMES, LOUISE B., LEARNED, JANET, METRAUX, RUTH W., & WALKER, R. N. *Rorschach responses in old age.* New York: Hoeber-Harper, 1954.

BECK, S. J. *Rorschach's test. III. Advances in interpretation.* New York: Grune & Stratton, 1952.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1957.

FREUD, S. *Basic writings of S. Freud.* New York: Random House, 1938.

HALL, C. S. A cognitive theory of dream symbols. *J. gen. Psychol.,* 1953, **48**, 169–186.

HALPERN, FLORENCE. *A clinical approach to children's Rorschachs.* New York: Grune & Stratton, 1953.

KLOPFER, B., AINSWORTH, MARY D., KLOPFER, W. G., & HOLT, R. R. *Developments in the Rorschach technique.* Vol. 1. New York: World Book, 1954.

LINDNER, R. M. The content analysis of the Rorschach protocol. In L. E. Abt & L. Bellak (Eds.), *Projective psychology.* New York: Knopf, 1952.

OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.,* 1952, **49**, 197–273.

PHILLIPS, L., & SMITH, J. G. *Rorschach interpretation: Advanced technique.* New York: Grune & Stratton, 1953.

PIOTROWSKI, Z. A. *Perceptanalysis.* New York: Macmillan, 1957.

RORSCHACH, H. *Psychodiagnostics.* New York: Grune & Stratton, 1943.

SARASON, S. B. *The clinical interaction.* New York: Harper, 1954.

SCHAFER, R. *Psychoanalytic interpretation in Rorschach testing.* New York: Grune & Stratton, 1954.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

A NOTE ON LEVIN'S REPORT OF FINDINGS WITH PHENOBARBITAL, PROMETHAZINE, CHLOR-PROMAZINE, AND PLACEBO

LLOYD K. SINES
University of Minnesota

In view of the many reports in the literature and other sources of information relative to designing and conducting adequate studies in the area of psychopharmacology (Kline, 1959; Rashkis & Smarr, 1958), there seems little excuse for the persistence of many of the major tactical and other errors apparent in this area of important and copious research. A recent study by Levin (1959), reported in this Journal, exemplifies several such unnecessary methodological errors. In view of the fact that reports in the professional literature of psychopharmacological studies will undoubtedly increase in the immediate future, continuing efforts must be made to improve research design and methodology and, also, to maintain high standards for the publication of such reports.

I should like to note two types of errors exemplified in Levin's study and, also, to comment on what appears to be an error in the author's interpretation of his own data.

Methodological Shortcomings

1. The use of an arbitrary and uniform dosage of each medication for all patients in the study constitutes the major methodological error in the study. This practice cannot be defended in view of the knowledge of individual differences among persons in their responsiveness to various chemical compounds. In this connection, it has been reported by Winkelman (1957) that a ratio of 35 to 1 has been found in the individual patient tolerance of, or reaction to, ataractic drugs. At the present time, the use of an experienced physician's clinical judgment in the determination and individualization of drug dosage on a patient-to-patient basis is much preferred over the practice of employing a single uniform dosage of the therapeutic agents used.

2. The possibility of bias in the behavioral ratings cannot be ruled out in view of the fact that the raters may have been aware of differences among the groups in the preparations which the patients received. That is, though it is stated that chlorpromazine and its placebo were "similar" in appearance, presumably the phenobarbital and promethazine preparations were distinguishable in appearance (and possibly in other ways) from the other two agents. Further, it is common knowledge that even under the so-called "double-blind" regime, ward personnel (and raters) are able to predict with high accuracy the actual group identification of patients receiving alternative therapeutic agents (Hall & Dunlap, 1955). It is, of course, not inevitable that such information leads to biased judgments of change among patients, although it is necessarily a limiting condition in the experimental design. The use of standard liquid or parenteral modes of drug administration is a desirable (though often impractical) procedure aimed toward minimizing possibilities for the identification of the specific group status of patients on such a priori grounds.

3. An over-all attrition rate of 23% among patients originally included in the study appears somewhat high and suggests the possibility that even if the original patient sample were a random or representative one with respect to the parent population, the final analyses (and therefore generalizations) may well have been based upon a unique or selective sample of subjects.

Statistical Weaknesses

1. Data (and statistical analyses) supporting the conclusion that fewer symptoms were observed in all groups at the end of the follow-up period than at the beginning of the study are not offered. The graph presented by the author, on the other hand, suggests that the phenobarbital and placebo groups were about the same (in terms of Wittenborn scores) at the end of the follow-up period as at the beginning of the study. The latter, of course, would imply that a general or overall decrease in symptomatology at the end of that period could be accounted for by improvements among chlorpromazine and promethazine group patients.

2. Similarly, data are not provided which presumably provided the basis for the conclusion that there was "a significant increase in symptoms" among the groups from the end of the treatment period to the end of the follow-up period. On the contrary, the graph provided suggests only *minimal* regression to initial level among patients in the chlorpromazine group. It is conceivable, therefore, that the therapeutic improvement in the latter group was maintained during the follow-up period.

3. In view of the (significant?) differences on the Wittenborn scale among the groups at the beginning of the study, the use of the analysis of covariance at the end of the treatment period and at the end of the follow-up period would have been preferable to the use of the analysis of variance.

4. The investigators failed to pursue the data sufficiently by neglecting the application of appropriate small sample statistics (t tests between related means) to the data. Such tests might well have revealed that, while all *drug* groups improved during the treatment period, placebo group patients remained essentially unimproved and, further, that all groups *except chlorpromazine* returned to their previous symptomatic levels after a two-month no-medication period.

Finally, a gross error in the interpretation of the data presented appears to have been made, i.e., differential effects among the experimental groups *were* demonstrated at the end of the treatment period ($F = 21.26$, $P < .01$), and may well have persisted through the follow-up phase (an analysis of variance was not carried out on the latter data), indicating that in addition to whatever general improvements were common to the groups, there was further *differential change* due to the several treatments. The author's conclusion that "all groups showed improvement of similar degrees" is, therefore, in direct contradiction to the data offered.

While this writer agrees with Levin's statement that "in designing drug research, great care must be taken to control incidental variables [and] caution must be used in interpreting uncontrolled clinical studies in which treatment methods are evaluated," I fail to see how the data presented by Levin constitute an empirical corroboration of such cliché-like admonitions.

I find such methodological bungling in psychopharmacological studies at this late date exasperating and, further, such tactical and interpretive errors inexcusable in articles published in the psychological literature.

Received May 1, 1959.
Early Publication.

REFERENCES

Hall, R. A., & Dunlap, D. J. A study of chlorpromazine: Methodology and results with chronic semi-disturbed schizophrenics. *J. Nerv. Ment. Dis.*, 1955, **122**, 301–314.

Kline, N. S. (Ed.) *Psychopharmacology frontiers.* Boston: Little, Brown, 1959.

Levin, M. L. A comparison of the effects of phenobarbital, promethazine, chlorpromazine, and placebo upon mental hospital patients. *J. consult. Psychol.*, 1959, **23**, 167–170.

Rashkis, H. A., & Smarr, E. R. A method for the control and evaluation of sociopsychological factors in pharmacological research. *Psychiat. res. Rep.*, 1958, **9**, 121–130.

Winkelman, N. W. An appraisal of chlorpromazine. General principles for administration of chlorpromazine, based on experience with 1,090 patients. *Amer. J. Psychiat.*, 1957, **113**, 961–971.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

A REPLY TO SINES' NOTE ON A COMPARISON OF THE EFFECTS OF PHENOBARBITAL, PROMETHAZINE, CHLORPROMAZINE, AND PLACEBO UPON MENTAL HOSPITAL PATIENTS

MONROE L. LEVIN
Oneida Consolidated Schools, Oneida, New York

Sines' recent note (Sines, 1959) criticizes aspects of the methodology and statistical methods used in the writer's recently published research (Levin, 1959). In evaluating the original paper in the light of Sines' criticisms and Sines' comments, attention should be paid to the initial paragraphs of the study; the major purpose of the study was to determine whether, as had been anticipated in advance, variables other than drugs were of sufficient strength to produce measurable and observable changes in the behavior of mental hospital patients, and whether the behavioral changes would be in the directions anticipated.

Methodology

Sines objected to use of arbitrary and uniform dosage in each of the subject groups on the grounds that there are wide fluctuations in individual responses to chemical compounds. He further reports that others have found a wide range of patient responses to ataractic drugs, concluding that individualized drug dosage is to be preferred to uniform dosage of therapeutic agents.

The study in question was conceived in 1955, at a time when ataractic drugs were in the process of becoming available in the state hospital at which the study was conducted. At that time, uniform dosages were often used, both because of the relative newness of the ataractic drugs in state mental hospitals, and because of the large numbers of patients for whom individual physicians were responsible.

Sines' criticism did not take into account the fact that if the ataractic used in the study does, as was then being reported, produce dramatic and highly positive results in an almost consistent way, use of the agent in even the limited fashion reported should have produced positive and differential results similar to ones reported earlier by others. It was recognized that uniform dosage might attenuate the results, but it was not believed that this would be sufficient to completely attenuate the effects of the drugs which were used.

Use of physician-directed, modifiable dosages of drugs and placebo was considered but necessarily rejected as a possibility because of the innumerable problems it would have presented both in the administrative structure of the hospital and in the design of the study.

The dosages of ataractic and antihistamine were suggested by the pharmaceutical concerns involved as ones which were then considered to be optimal under the conditions of the research.

Sines states that bias of raters may have been a major factor in slanting the findings, because of the ability of ward personnel to identify the various drugs and placebo used. Attention is called to the methods used to minimize bias. If any of the ward personnel were biased because of their ability to distinguish between the various tablets, the *identical* (not "similar" as Sines states) appearance of the placebo and chlorpromazine might well have exaggerated, rather than minimized differences between the groups, and might well have favored a greater difference than was obtained between the ataractic, placebo, and other drugs.

Since patients were distributed in almost random fashion throughout the hospital, it was assumed that any bias which arose would be inconsistent and sufficiently attenuated to become negligible.

Sines' third criticism is of the attrition rate (23%) reported. He indicated that the rate of attrition may have been a reflection of sampling bias. Every patient in a 2,000 bed hospital who met the selection criteria was initially included in the study. Reference is made to the original paper in which it was indicated that most subjects were lost for administrative reasons. "Administrative reasons" included such factors as failure of ward personnel to administer drugs, failure to rate patient behavior at appropriate intervals, and failure to administer appropriate drugs. The vast majority of subjects were lost because of administrative reasons, as was originally reported. The attrition rate, therefore, was clearly not due to sampling factors but, rather, to factors which have no demonstrable relation to patient behavior or grouping in the study.

Since the hospital at which the study was conducted serves an area representing one quarter of the geographic area of the state in which it is located, and since almost all eligible patients participated, it is doubtful that the sample represented anything other than that which it was intended to.

Statistical Method

In view of the purpose of the study, the data were examined to determine if there were differential shifts over time in symptoms, as measured by means of the Wittenborn scales. Differential shifts should produce significant interaction between treatment and time. The analysis of variance (see Table 1, Levin, 1959) clearly showed that there was no significant interaction.

Sines appears to have confused the meaning of the analysis of variance reported. The significant F reported in Table 1 does *not* indicate, as he says, that there were differential effects among the experimental groups at the end of treatment period; the F merely indicates that significant shifts occurred in the symptom scores of *all* groups during the course of the study.

Received June 8, 1959.

REFERENCES

LEVIN, M. L. A comparison of the effects of phenobarbital, promethazine, chlorpromazine, and placebo upon mental hospital patients. *J. consult. Psychol.*, 1959, 23, 167–170.

SINES, L. K. A note on Levin's report of findings with phenobarbital, promethazine, chlorpromazine, and placebo. *J. consult. Psychol.*, 1959, 23, 461–462.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

ABSENCE OF ACQUIESCENCE RESPONSE SET IN THE TAYLOR MANIFEST ANXIETY SCALE[1]

LOREN J. CHAPMAN
University of Chicago

AND DONALD T. CAMPBELL
Northwestern University

When attitude and personality scales of the agree-disagree or true-false format contain unequal numbers of items worded in the direction of the two ends of the personality dimension, the response to the content of the items may be confounded with general tendencies to agree or disagree (Cronbach, 1946). A number of recent studies have investigated such acquiescence response set in the F scale from *The Authoritarian Personality* study (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) by means of correlating scales or reversed items with scales of original items. A recent review (Chapman & Bock, 1958) has shown that all of these studies found evidence of substantial acquiescence variance.

Another scale which might be suspected of acquiescence bias is the Taylor Manifest Anxiety Scale (Taylor, 1953), in which 39 of the 50 items are worded so that agreement with the item contributes toward a high MA score.

A reversed MAS was constructed by simple literal reversals of meaning, e.g., for "I am often sick to my stomach," the reversal was "I am seldom sick to my stomach."

If there were acquiescence variance in the MAS, it would be expected to influence the correlation of the MAS with other measures which themselves are correlated with acquiescent tendencies. To study such possible effects, each *S* was also given original and reversed F and E scales (Adorno et al., 1950).

Two versions of the test were prepared, each having half of the items from each scale in positive form and the other half negative, with each item appearing in its positive form in one scale and its negative form in the other. Each *S* received the two forms one week apart. Since the E scale measures attitudes toward minority groups, Jewish, Negro, and Oriental *S*s were not included. The final experimental group was 184 college students.

As an additional intelligence measure for 86 of the 184 *S*s, OSPE percentile ranks were obtained from the university admission's records, and for another 70 a sum of verbal plus mathematical scores on the college boards was employed.

Results

The Kuder-Richardson reliabilities of the positive and negative forms of the MAS were .87 and .85, respectively, and the correlation between positive and negative halves was .84. The correlation is seen to be almost as high as the reliabilities of the scales, a finding which is in striking contrast to the results of similar studies on the F scale, including that of Christie, Havel, and Seidenberg (1958). We conclude that the items of the MAS, unlike those of the F scale, are subject to little or no acquiescence bias. This is probably attributable to the quite specific and personal reference of the items.

Table 1 shows the intercorrelations of the two MA scales with the other measures. Note that both the positive and negative forms of MAS show correlations with the original F scale (F_{pos}), significant beyond the .01 level of .19. No significant correlation with intelligence or with the E scale is found.

Note that neither MA_{pos} nor MA_{neg} represents the original MAS. When the MAS is re-

---

[1] This study was supported in part by the Social Science Research Committee of the University of Chicago and in part by the Graduate School, Northwestern University.

Table 1

Relationships of the MA Scales to Other Measures[a]

	F_{pos}	F_{neg}	E_{pos}	E_{neg}	O.S.P.E.	V.+M. Col. Brds.
MA_{pos}	$\underline{.32}$	.07	.11	.04	.00	−.10
MA_{neg}	$\underline{.21}$	.06	.05	.08	−.01	−.09

[a] Underlined values are significant at the .01 level.

constituted from the item pool of the present study, it has a reliability of .84. The correlation between MAS and F_{pos} is .29.

Received July 23, 1958.

REFERENCES

Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.

Chapman, L. J., & Bock, R. D. Components of variance due to acquiescence and content in the F scale measure of authoritarianism. *Psychol. Bull.*, 1958, **55**, 328–333.

Christie, R., Havel, Joan, & Seidenberg, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

Cronbach, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475–494.

Taylor, Janet A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

BRIEF REPORTS

VALIDATION OF DOPPELT'S WAIS SHORT FORM WITH A CLINICAL POPULATION

HUGH CLAYTON AND DAN PAYNE
Colorado State Hospital

In 1956, Jerome Dopplet published his findings relative to a brief form of the WAIS. The brief form consisted of the following four subtests: Arithmetic, Vocabulary, Block Design, and Picture Arrangement. The sum of the four scaled scores is multiplied by 2.5 and increased by a constant which varies with the subject's age. The constants, for their respective age ranges, are as follows: 16–34: 10, 35–44: 9, 45–54: 8, 55–64: 7, 65–74: 5, 75+: 4. The final sum is equivalent to the Full Scale "Scaled Score." The IQ is then found in the IQ table appropriate to the subject's age.

Using the above formula, Doppelt computed the Brief Scale Scores for Wechsler's standardization population, broke them up into age groups, and then ran Pearson correlations between the Brief Scale IQ and the Full Scale IQ for each of the age groups. Pearson *r*'s ranged from .95 to .96.

Wechsler's standardization population consisted of "normal" subjects. The question raised here is that of validity with a clinical population. There would be a real possibility that the correlations might not run nearly so high with a population of mentally disturbed patients whose intellectual functioning is erratic in varying degrees. Another question to be answered concerns the possibility that the Brief Scale IQ might be more valid with certain clinical groups than with others.

All active test files in the psychology department of the hospital were pulled and examined. Those whose scores were based on all 11 subtests of the WAIS were accepted as subjects. The files were then classified by the following diagnostic categories (based on hospital staff decisions): 1. Schizophrenia ($N = 72$); 2. Personality disorders ($N = 51$); 3. Organics ($N = 27$); 4. Mental defectives ($N = 28$); 5. Mixed diagnoses ($N = 50$). Pearson correlations were then run between the estimated IQs and the Full Scale IQs, separately for each of the first four categories, and then together (including category "5").

The *t* tests failed to reveal any significant differences between the means of estimated IQs and the means of Full Scale IQs.

The obtained correlations for the first four groups and for the total are as follows: .97 for schizophrenics; .96 for personality disorders; .96 for organics; .92 for mental defectives; and .97 for the total. Standard errors of estimate ranged between 3.4 and 4.6. A *t* test for the significance of the difference between the *r* of .92 for mental defectives and that of the other three groups proved significant at the .02 level. With the possible exception of mental defectives, then, the brief form of Doppelt appears as valid an estimate of the Full Scale Score with a clinical population as it was with the "normal" standardization population.

Brief Report.
Received March 13, 1959.

REFERENCE

DOPPELT, J. E. Estimating the full scale score on the Wechsler Adult Intelligence Scale from scores on four subtests. *J. consult. Psychol.,* 1956, **20**, 63–66.

---

[1] An extended report of this study may be obtained without charge from Hugh Clayton, Psychology Department, Colorado State Hospital, Pueblo, Colorado, or for a fee from the American Documentation Institute. Order Document No. 5922, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

THE VALIDITY OF A SELF REPORT MEASURE OF ANXIETY AS A FUNCTION OF THE TIME INTERVAL COVERED BY THE INSTRUCTIONS[1]

BARCLAY MARTIN
University of Wisconsin

The instructions accompanying most paper-and-pencil scales of personality are vague with respect to the time interval for which the person is rating himself, although the usual implication is that the person is answering the items in terms of how he has generally been during his life. For many traits, and especially for a trait such as anxiety, there is little doubt but that most people vary considerably from day to day or month to month in the extent to which they possess the trait. It is not surprising therefore to find that a scale such as the Taylor Manifest Anxiety Scale, taken in terms of how anxious one generally is, frequently fails to predict anxiety as assessed by other methods in particular situations. The purpose of the present research was to experimentally explore the effect of different instructional time periods on the relationship of a self-report measure of anxiety to a criterion assessment of anxiety in a specific stress situation.

Anxiety was psychometrically measured by the Feeling Inventory, a specially constructed forced-choice inventory, consisting of 23 triplets of adjectives selected on the basis of a previous internal consistency item analysis, with adjectives such as "tense" and "relaxed" making up the respective ends of the scale.

*S*s were first seen in an individual stress situation in which they were confronted in close proximity by two experimenters who rather obviously made continuous ratings on the *S*s throughout the 15-minute session. *S*s were told that they were going to be asked to do things that would allow the experimenters to understand what they were really like as persons, what their inner wishes and fantasies were, etc. *S*s were then asked to respond to one Rorschach card, one TAT card, and to free associate for a short period. The two experimenters independently tallied nervous movements throughout the session and made a global rating of anxiety on a seven-point scale at the end of the session. Continuous GSR recordings were also obtained.

At the end of the session, *S*s were randomly assigned to one of three groups and given instructions to take the Feeling Inventory in terms of (*a*) how they generally feel, (*b*) how they have felt during the last month, and (*c*) how they just felt in the 15-minute session.

Only for the "15-minute" instruction group was there a significant tendency for Feeling Inventory scores to correlate with any of the criterion anxiety measures in the expected direction. The highest correlation for this group was .44 for base level skin conductance. For the "last month" instruction group, there were unexpected negative correlations between rated anxiety and nervous movements, and the Feeling Inventory. Taylor Manifest Anxiety scores obtained in an earlier group session did not correlate significantly with any of the criterion measures.

Brief Report.
Received January 26, 1959.

---

[1] An extended report of this study may be obtained without charge from Barclay Martin, Psychology Department, University of Wisconsin, Madison, Wis., or for a fee from the American Documentation Institute. Order Document No. 6017, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

MANIFEST ANXIETY AND PERFORMANCE ON PROBLEM SOLVING TASKS[1]

DURGANAND SINHA[2]
Indian Institute of Technology, Kharagpur, India

AND TRIPAT RAJ SINGH[3]
Patna University

In this study, the performance of 20 high scoring and 20 low scoring subjects on the Taylor Scale of Manifest Anxiety was compared on a set of six problem solving tasks which were novel and of moderate difficulty. Three of the tasks were sensorimotor in character (Substitution Test, Design Sorting Test, and Katona Matchstick Problems), and the other three involved largely mental activities (Line-Pursuit Test, Mixed Sentences, and Number Series). Within the general framework of Hull's behavior theory, it was expected that tasks being complex involving a hierarchy of competing responses, the higher drive level, as measured by high scores on the Taylor scale, would lead to impairment of performance.

Though HA *Ss* had committed a few more errors on each task, the performance of the two groups did not differ significantly in respect of errors. However, on all the Katona problems more LA than HA *Ss* had solved them, χ^2 being significant in all instances (Problem 1: $\chi^2 = 6.66$, $p = .01$; Problem 2: $\chi^2 = 5.23$, $p = .02$; Problem 3: $\chi^2 = 3.95$, $p = .05$).

But the HA group consistently required more time for solving both the sensorimotor and mental tasks, t ratios being significant in all cases but one (Sorting: $t = 4.0$, $p = .01$; Substitution: $t = 3.4$, $p = .01$; Katona: $t = 3.4$, $p = .01$; Line-Pursuit: $t = 3.2$, $p = .01$; Mixed Sentences: $t = 3.88$, $p = .01$; and Number Series: $t = 1.59$, $p =$ not significant).

The inferiority of HA *Ss* can be explained in terms of difficulty of integration of responses and the selection of the correct one out of many possible responses. Increased drive level multiplied indiscriminately the strength of all competing responses. Irrelevant and wrong movements were made and had to be corrected. This interference showed itself in slightly higher error score and significantly longer time required for the solution. Performance was of less integrative character due to "anxiety-produced interference" (Malmo & Amsel, 1948) or general "disorganization of behavior" (Davis, 1948). The effect on error score was masked off, because *Ss* were allowed to correct themselves during performance, and only those errors that persisted till the end were recorded. However, error-corrections required time, and consequently the time scores in high anxiety *Ss* were raised.

Relatively longer time required by HA *Ss* in solving problems has obvious implication on results obtained on tests where speed of performance is a significant consideration. Anxiety level seems to become a significant variable in test performance. It is particularly so on speed tests.

Received April 6, 1959.
Brief Report.

REFERENCES

Davis, D. R. Increase in strength of a secondary drive as a cause of disorganization. *Quart. J. exp. Psychol.,* 1948, **1**, 22–28.

Malmo, R. B., & Amsel, A. Anxiety-produced interference in serial rote learning with observations on rote learning after partial frontal lobectomy. *J. exp. Psychol.,* 1948, **38**, 440–454.

---

[1] An extended report of this study may be obtained without charge from Durganand Sinha, Indian Institute of Technology, Kharagpur, India, or for a fee from the American Documentation Institute. Order Document No. 6018, remitting $1.25 for microfilm or $1.25 for photocopies.

[2] Formerly at the Institute of Psychological Research & Service, Patna University, India, where this study was conducted.

[3] A graduate student in Psychology at Patna University during the sessions 1955–57.

Journal of Consulting Psychology
Vol. 23, No. 5, 1959

PSYCHOLOGICAL TEST REVIEWS

Caligor, Leopold. *The Eight Card Redrawing Test (8CRT)*. Kit containing test instructions, blanks, score sheets, and scoring masks, pp. 148. 8CRT, P.O. Box No. 31, Gracie Station, New York 28, New York. Manual: A New Approach to Figure Drawing. Springfield, Ill.: Charles C Thomas, 1957.

This technique represents a modification and elaboration to the Draw-A-Person test. The subject is requested to "Make a picture of a whole person." When the drawing is completed a sheet of transparent onionskin paper is placed over it and the subject is asked to make a second drawing on this onionskin. His exact instructions are as follows: "You may do anything you like with this picture of a person. You may add to it, take away from it, change it or leave it alone. Only again, make a picture of a whole person." When the second drawing is completed, the first drawing is then concealed and an onionskin sheet placed over the second drawing. A third drawing is requested from the subject and he receives the same instructions as he did for the second drawing. This process is continued until a total of eight drawings have been obtained.

Caligor devised this technique because he felt that "The single drawing left so many questions unanswered." He expected that multiple drawings would provide a much richer yield of data about the subject, particularly in terms of the continuities and discontinuities that were apparent in the total series. The manual for the 8CRT contains an elaborate scoring system for evaluating well over twenty different variables. These variables relate primarily to structural aspects of the drawings. Some of the principal scoring categories are of the following order:

Degree to which complete figure drawn
Height of figure
Placement of figure on page (top, bottom, right, left)
Stance of figure
Ratio of head size to body size (height & width)
Sex of figure
Transparencies
Erasures
Relation of body walls and body areas of the figure to the preceeding figure
Degree of movement of figure
Manner in which the body is filled in (e.g., amount of detail and shading)
Line quality (tremble, pressure, thickness)
Symmetry of placement of body parts
Clothing
Omission of body parts

Each category is interpreted in terms of the degree to which it shifts and varies over the entire series of eight drawings. Caligor offers brief formulations concerning what he considers to be the meanings of the various categories. Illustratively, in referring to "placement on the page" he says, "Placements other than center reveal limitation in the ability to adequately cope with the environment. They indicate inadequate cognizance or handling of stimulation and attempted avoidance of the full impact of the environment." Further, in his description of the significance of "erasures" he states, "Most people erase on two or more drawings. The absence of erasures . . . reflects limitation of corrective and critical faculties, possible lowered reality testing and frequently a labile reaction to stimulation. Excessive erasures reflect hypercriticality, blocking or conflict and overcontrolled resistance to stimulation." In order to clarify how his scoring categories may be applied he presents detailed analyses in his manual of the drawings of three different subjects. These analyses are of a broadly impressionistic character and apparently involve the intuitive discovery of a meaningful pattern in a large array of scores.

Caligor does refer to three studies which were undertaken for validation purposes, but his descriptions of these studies are extremely brief and it is clear that he did not attempt to test directly the validity of the meanings assigned to the various scoring categories. One can see that his formulations regarding the significance of given variables are based almost entirely on his clinical experiences and that one must accept them on faith. The 8CRT is presently a collection of hunches that have been formalized by Caligor into a scoring system. Some of his hunches are novel and interesting and may prove eventually to have an enriching effect upon figure drawing analysis. Thus one is particularly impressed with the novelty and potential importance of his emphasis on such factors as spatial directionality (e.g., up vs. down and right vs. left), symmetry, and mode of maintaining continuity from one drawing to another.

Overall, though, it must be said that his scoring categories are not bound together by a unifying concept or viewpoint. They seem, on the contrary, to consist of heterogenous "signs" which were assembled in an arbitrary fashion. One must also question the value of Caligor's mode of defining many of the scoring categories. His definitions are often very vague and hazy. Illustratively, he refers to various signs as indicating "anxiety," "conflict," "immaturity," "lowered ability to orient oneself in the environment," "ability to use inner resources," "awareness of objects or other persons in the environment." How much more would one know about a given subject for having acquired such vague bits of information about him? The 8CRT does indeed present some new ideas about figure drawing analysis, but it lacks the rationale or validation to be considered a formal test.—*Seymour Fisher.*

Bennett, George K., Seashore, Harold G., & Wesman, Alexander G. *Differential Aptitude Tests.* Manual (3rd ed.), pp. 94. New York: Psychological Corp., 1959.

The new manual follows very closely the format of the previous (1952) edition (see *J. consult. Psychol.*, 1953, **17**, 78). The main change is that a 16-page research supplement has been added. This supplement offers data designed to validate the combination of the *Verbal Reasoning* and *Numerical Ability* scores into an index of scholastic aptitude and, in addition, six other studies selected for their special interest to users of the tests. The bibliography has increased from 27 to 105 items. The manual remains a model of its category.—*E. S. B.*

Carter, H. D. *California Study Methods Survey.* Grades 7 to 13. Test booklet (no time limit) with manual, 16 pp., answer sheets, scoring stencils. Los Angeles: California Test Bureau, 1958.

This 150-item inventory yields four scores, Attitudes toward School, Mechanics of Study, Planning and System, Verification. The last one based on items with "the thirty most popular responses" is supposed to validate the other scales. A good deal of work seems to have gone into the development of this instrument; its reliability seems adequate, and there is encouraging evidence of its validity for use in educational diagnosis with high school students. Little evidence is offered to support the supposed function of the Verification score.—*E. S. B.*

Edwards, Allen L. *Edwards Personal Preference Schedule.* Manual, Revised 1959, pp. 27. New York: Psychological Corp., 1959.

This revision of the 1954 manual (see *J. consult. Psychol.*, 1956, **20**, 321–324) differs in only minor respects from the original. The directions for administration now include a section on machine scoring; a general adult sample now provides an additional norm group; the bibliography has been expanded from 9 to 82 references, most of the new references representing studies in which the EPPS figured. The general adult sample is less than adequate for normative purposes because it consists of a nationwide sample of male and female household heads who are members of a consumer purchase panel used for market surveys. It is not clear what this means with regard to its possible biases with respect to age, education, and class status. The mean scores of this sample vary from those in the other norm group, college students. But there is no effort to analyze the significance of these differences for the validity of the test. In fact, it is astonishing that no effort was made to summarize the relevance of the large body of studies in the bibliography to the test's validity and reliability. The reviews of the first manual pointed to its deficiencies with regard to validating evidence. One wonders why the author and publisher bothered to offer a revision which revises so little and is so unresponsive to criticsm.—*E. S. B.*

Mitchell, C. *California Analogies and Reasoning Test.* High School and College. Forms A and B. Test booklet, with manual, 16 pp., answer sheets, scoring stencils. Los Angeles: California Test Bureau, 1958.

A scholastic aptitude test in which content drawn from natural science, mathematics, social science, English, and literature has been cast into analogy form. The manual offers the necessary information on reliability norms, directions for administration and scoring, etc. The studies offered to support its predictive validity are less than adequate because they are based on samples tested either concurrently with the collection of criterion data (academic grades) or where the grades had already been earned. Thus, there is evidence to support the test's value for concurrent validity or for postdiction, but none for prediction.—*E. S. B.*

Stone, Joics B. *S-O Rorschach Test.* College and adult. Test booklet, with manual, pp. 58, answer sheet, scoring stencils. Los Angeles: California Test Bureau, 1958.

The S-O (Structured-Objective) Rorschach Test, as its title suggests, attempts to tap the same psychological processes as its namesake via modified technique designed for group or self administration to yield a uniform number of responses and an objective method of scoring. For each blot, 30 percepts are selected and grouped into 10 triads so as to provide a basis for scoring on location, determinants, limited content factors (human and animal), and frequency (modal and rare). The 10 original stimuli are presented in one of three ways: projected on a screen, individual inkblot card sets, or via an illustrated booklet. The subject is asked to choose the one item in each triad that is most clearly represented by the blot or by some part of the blot. Choice of an item in a triad contributes to each of the four general categories of scoring much as in the standard situation.

Standard scores are provided for 15 variables based upon a normative sample which tries to make up in numbers what it lacks in definitiveness. The normative sample is an amalgam of 967 Brigham Young University students and 7094 selected occupational groups. The reported test-retest reliabilities (one-week interval) are acceptable but not great, ranging between .62 and .90. The 15 factors are either taken singly or combined in various ways to derive 26 different attributes.

The validity of the test seems to rest mainly on two bases: first, its supposed relationship to Rorschach phenomena; second, two empirical studies. Without considerable empirical evidence, there is great room for doubt that responses to the SORT are mainly tapping perceptual phenomena analogous to its prototype. The method of test administration contains no procedures that ensure that the subject has seen the percept that he chooses nor for that matter that he has even looked at the blot. Response sets of various sorts, particularly social desirability stereotypes, are probably greatly enhanced under such circumstances. The manual gives a brief report of two unpublished validity studies. In one, the SORT increased the correlation of high school grade point average with first year grades from .59 to .68, using the best 2 of 15 scores. No cross-validation data are offered. In the other, the SORT variables are correlated with supervisors' ratings in 29 occupational groups. A suggestive but not impresive array of correlations was obtained. Again no cross-validation.

This all sums up to the fact that we have here an interesting new experiment in adapting Rorschach testing techniques to the need for large scale testing and objective scoring and interpretation. But it is still *experimental,* and every effort should be made to warn against adopting it for operational use. I do not believe the "Preliminary Edition" in the title or the caution section in the manual, largely irrelevant to this issue, represent sufficient effort to emphasize the fact that this instrument is not ready for operational use.—*E. S. B.*

JOURNAL OF CONSULTING PSYCHOLOGY

Vol. 23, No. 6

December 1959

Contents

ARTICLES

Mental Health Consultation: Theory and Practice: *Arthur J. Bindman* - - - - - - 473
The Relative Contribution of Four Kinds of Data To Accuracy in Personality Assessment: *Lloyd K. Sines* - 483
A Note on the Postural Sway Distribution: *Fred Heilizer* - - - - - - - - - - - 493
The Set to Claim Undesirable Symptoms in Personality Inventories: *Clinton B. DeSoto and James L. Kuethe* - - - - - - - - - - - - - - - - - - 496
Effect of a Training School Experience on the Personality of Delinquent Boys: *Sylvan B. Caditz* - 501
The Effect of Item Style on Social Desirability and Frequency of Endorsement: *Arnold H. Buss* - 510
Social Desirability and MMPI Performance: The Individual Case: *James Bentley Taylor* 514
The Attitudes of Mothers of Juvenile Delinquents Toward Child Rearing: *Jeff M. Madoff* - 518
Sexual Symbolism, Personality Integration, and Intellectual Functioning: *Rudolph Moos and Paul Mussen* - - - - - - - - - - - - - - - - - - 521
Weight Problems and Psychological Factors: *Joseph Friedman* - - - - - - - - - 524
Social Desirability and the Expression of Aggression under Varying Conditions of Frustration: *Joel Allison and David E. Hunt* - - - - - - - - - - - - - - 528
Correlates of Attitudes Among Clinical Psychology Students: *Sheldon R. Roen* - - - 533
Requisite Conditions for Basic Personality Change: *Albert Ellis* - - - - - - - - 538
A Note on an Index for the Comparison of Profiles on the "Circle": *William L. Hays* 541
The Sentence Completion Test as a Predictor of Continuation in Psychotherapy: *E. Wesley Hiler* - 544
A Factor Study of the MMPI Psychopathic Deviate Scale: *Alexander W. Astin* - - 550
A New Maximum CA for the Draw-A-Man Test: *Anna Carr Mitchell* - - - - - - 555
A Note on "Social Desirability As a Variable in the Edwards Personal Preference Schedule": *Allen L. Edwards, Calvin E. Wright, and Clifford E. Lunneborg* - - - 558

BRIEF REPORTS

Bender-Gestalt Rotations and EEG Abnormalities in Children: *Sherwood B. Chorost, George Spivack, and Murray Levine* - - - - - - - - - - - - - - - - - - 559
Eye-Ear Emphasis in the DAP as Indicating Ideas of Reference: *Albert V. Griffith and D. A. R. Peyman* - 560
The Validity of the Columbia Mental Maturity Scale: *Betsy Worth Estes, Frank Kodman, and Macy Akel* - 561
Personality Factors in Nursery School Children as Derived from Teacher Ratings: *Donald R. Peterson, Raymond B. Cattell, and Herbert C. Quay* - - - - - - - - 562
PSYCHOLOGICAL TEST REVIEWS - 563

(Volume Title Page and Contents Appear at the End of this Issue)

Brief Reports

The *Journal of Consulting Psychology* will accept Brief Reports of research studies in clinical psychology for early publication without expense to the author. The procedure is intended to permit the publication of soundly designed studies of specialized interest or limited importance which cannot now be accepted because of lack of space. Several pages in each issue will be devoted to Brief Reports, published in the order of their receipt without respect to the dates of receipt of the regular articles. Most Brief Reports appear in the first or second issue to go to press following their final acceptance.

An author who wishes to submit a Brief Report:

1. Sends the Brief Report, limited to one printed page and prepared according to the specifications given below.
2. Also sends to the Editor a full report of the research study, in sufficient detail to give a clear account of its background, procedure, results, and conclusions, which will be filed with the American Documentation Institute to insure indefinite availability.
3. Prepares at least 100 mimeographed copies of the full report, which the author will send without charge to all who request it as long as the supply lasts.
4. Agrees not to submit the full report to another journal of general circulation.

Specifications

Brief Report. The Brief Report should give a clear, condensed summary of the procedure of the study and as full an account of the results as space permits.

To insure that the Brief Report will be no longer than one printed page, its typescript, including all matter except the title and the author's lines, must not exceed 85 lines averaging 42 characters and spaces in length. Set the typewriter margins for short lines of 42 characters, which are 3.5 inches long in elite typing, and 4.2 inches long in pica.

The manuscript of the Brief Report must be *double spaced* throughout. Except for its short lines, it follows the standard style of the 1957 revision of the *APA Publication Manual*. Headings, tables, and references are avoided or, if essential, must be counted in the 85 lines. Each Brief Report must be accompanied by a footnote in the style below, which is typed on a separate sheet and *not* counted in the 85-line quota:

> [1] An extended report of this study may be obtained without charge from John Doe, 300 Market St., Prospect 6, Mass. (giving the author's full name and address), or for a fee from the American Documentation Institute. Order Document No. ——, remitting $—— for microfilm or $—— for photocopies.

Extended report. Because the extended report is intended for photoduplication, and is not copy to be sent to a printer, its style should differ in several ways from that of other manuscripts: (*a*) The extended report should be typed with single spacing for economy in duplication. (*b*) Tables and figures should be placed adjacent to the text which refers to them. A caption should be typed below each figure. (*c*) Footnotes should be typed at the bottom of the page on which reference is made to them. In other respects, the full report is prepared in the style specified by the *Publication Manual*.

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

MENTAL HEALTH CONSULTATION: THEORY AND PRACTICE

ARTHUR J. BINDMAN[1]
Massachusetts Department of Mental Health

Since 1952, the Division of Mental Hygiene of the Massachusetts Department of Mental Health has been developing a program of mental health consultation which is considered to be one of the primary services of its community mental health centers (Hallock & Vaughan, 1956; Vaughan, 1955). In general, mental health consultation is carried out by a new fourth member of the traditional three-man child guidance clinic team, whose basic role is to develop relationships with those key persons in the social system who are primarily concerned with families and children. The major focus of this work has been with teachers, although mental health consultation services have been provided to juvenile courts, probation officers, child welfare agencies, and many others.

The role of the consultant is varied. He is interested in helping key persons on a crisis-oriented basis in order to aid them in more effective utilization of their professional roles with regard to a particular child or client; secondly, he acts as a technical assistant in helping a consultee or agency make a better clinical appraisal or referral for those children who need specific diagnostic evaluation or follow-up (Caplan: 1954a, 1954b, 1956a, 1959; Hallock & Vaughan, 1956; Vaughan, 1955). The consultant is clinically trained and proficient, and he may be either a psychiatrist, psychiatric social worker, or a clinical psychologist who has received special training in mental health consultation.

Features of Mental Health Consultation

A definition of mental health consultation can be distilled out of the writings of a number of workers in this field (Caplan, 1956b; Prosser, 1957; Towle, 1950), as follows:

> Mental health consultation is an interaction process or interpersonal relationship that takes place between two professional workers, the consultant and the consultee, in which one worker, the consultant, attempts to assist the other worker, the consultee, solve a mental health problem of a client or clients, within the framework of the consultee's usual professional functioning. The process of consultation depends upon the communication of knowledge, skills, and attitudes through this relationship, and therefore, is dependent upon the degree of emotional and intellectual involvement of the two workers. A secondary goal of this process is one of education, so that the consultee can learn to handle similar cases in the future in a more effective fashion, and thus enhance his professional skills.

One can immediately see that similarities arise between consultation and the educational process. Caplan (1959) points out that consultation makes use of so-called opportunistic characteristics in the consultant-consultee relationship, while education generally emphasizes a more systematic approach, with predetermined curriculum and a planned learning exercise. Perhaps problem centered seminars and individual and group supervision come closer to the consultation process.

Other differences are noted between consultation and the processes of education, psychotherapy, and administration (Caplan, 1959; Prosser, 1957).

1. The consultee usually initiates consultation in regard to a current mental health or work problem, while in education or supervision the relationship is primarily initiated "from above." Thus, consultation is based for its acceptance upon the "authority of ideas" (Baker & Siegel, 1953), and not the authority inherent in administrative relationships or as in the teacher-pupil relationship.

---

[1] The author wishes to express his thanks to Warren T. Vaughan and B. R. Hutcheson, past and present Director of the Division of Mental Hygiene of the Massachusetts Department of Mental Health, for providing the opportunities to develop the mental health consultant role; to Gerald Caplan of the Harvard School of Public Health for providing the conceptual framework in the in-service training program seminars on consultation; and to his fellow mental health consultants with whom many hours of discussion concerning consultation have been shared.

2. The consultant usually enters the consultee's social system or field of forces on a temporary, short-run basis, and therefore the scope of his work is somewhat limited and usually concerned with a specific problem area. Educational methods may cover a wide range of problems and they may maintain teacher and pupil together for a fairly lengthy stay. In this way, consultation also differs from most types of psychotherapy in which patient and therapist usually develop a long-term relationship and the therapist is concerned with many aspects of problems within his patient.

3. In many cases the consultant will be of a different professional background from the consultee and not just a more senior person in the same field. In supervision this is often not true, particularly in casework. This may appear to be a minor distinction, but it serves to highlight the fact that case consultation attempts to utilize mental health principles whereby workers who do not consciously feel that they are directly involved in the mental health field can be helped to become more sensitive and secure in solving problems which hinder good mental health practices.

4. The consultant usually plays an advisory role, and does not have to implement a plan to solve the problem himself. The plan for the solution of the problem is left to the consultee. The latter may or may not wish to follow through, but regardless of his wish, he should not expect the consultant to carry out the steps of problem solution himself. He can, however, expect continued support and assistance from the consultant as the consultee tries to solve the problem. In this way consultation differs to some extent from formal education, especially when the teacher may actually observe a problem, solve it, and give the solution ready-made to the student. In administration, the administrator may be expected to learn about the problem in a relationship with a subordinate, and at the same time effect change by promulgation of policy or through directives.

5. Finally, the consultant is usually not in a position of administrative control over the consultee. The consultee is usually not held accountable for following through the results of consultation. All along the line, it must be clear that the consultee can "take it or leave it," whether this applies to initiation of the consultation or to the follow-up of suggestions. Baker and Siegel point out that consultation is more effective if the consultee uses his own competence to recognize his need for help and requests it. They further state that sometimes the consultee may seek consultation, not because he sees a need for it, but because of administrative pressure, or anxiety, or a feeling that this is the thing to do. In the long run, they suggest that asking for help may actually connote inadequacy on the part of the consultee (Baker & Siegel, 1953). Thus an adequate consultation and outcome are solely dependent upon the development of a positive relationship between consultant and consultee, with both having the option to withdraw. Mental health case consultation can be subdivided into three categories: (*a*) case-insight, (*b*) action-help, and (*c*) consultee crisis (Caplan: 1956a, 1959).

Case Insight Consultation

In this situation, the problem is one of a lack of insight by the consultee into the problems that surround the client. The client's behavior may not be understood, and this may result in an increasing appearance of maladapativeness to a situation on the part of the client. The consultee may become increasingly concerned with this behavior, but not know how he can handle the problem. The role of the consultant in this type of situation is to increase the understanding of the consultee by pointing out the basis for the psychological functioning of the client, but in a way which is appropriate to the consultee's professional background. Perhaps the best technique is the use of joint observation, since the consultant and consultee can observe and discuss their perceptions in a natural, nonthreatening fashion, with the consultee gaining new knowledge and insight into the client's behavior, which he can use to enhance the environment and reduce the disturbing behavior pattern of the client.

Action-Help Consultation

This type of consultation is used when the consultee lacks the skill or facilities in his agency to assist a client, nor has he the knowledge or method to call in outside agencies that could help for this purpose. The consultant, in an active fashion, evaluates the facilities and strengths of the agency and the community which might help the client. However, he works with the consultee in a somewhat indirect and nondirective fashion to help the consultee feel that his views are clarified and supported, so that he can work out his plans in as able a fashion as possible without the consultant's plan being superimposed in a direct manner. Usually, through this method, the consultee's horizons begin to broaden and he becomes more flexible and able to handle more difficult problems without as much consultation assistance in the future. As time goes on the consultant can also give more information of a direct nature concerning experiences of other consultees with similar problems, which may enable the consultee to broaden his armamentarium of techniques for assisting clients. In the long run, this experience should help the consultee become a more sensitive case-finding and referral agent who can use facilities to a better advantage.

Consultee Crisis Consultation

In a number of cases, a disturbed or disturbing situation in a client may trigger off a crisis in a consultee (Caplan, 1956b; Vaughan, 1955). Quite often it is a type of behavior in the client which stimulates the re-emergence of similar problems in the consultee which he has learned to control or suppress. On the other hand, a crisis in a consultee may precede a crisis in a client, and in turn trigger off disturbing behavior in the client. In either case, the consultee begins to distort and stereotype his relationship with the client, and begins to project his own past difficulties into this relationship. He may feel that the client has a problem, when in truth the client serves as a displacement object for the consultee's disturbance. This often causes a great reduction in the consultee's ability to function in a smooth, effective fashion in his professional role. Therefore one of the primary aims of mental health consultee crisis consultation is suggested by Caplan (1956b, p. 77); namely,

> . . . improving educators' use of the self in the professional context, so that their personality strengths may be more effectively mobilized for the mental health of their pupils, and so that their unsolved emotional problems may not interfere with the utilization of their existing professional knowledge and skills.

The role of the consultant in this type of situation is to develop first a supporting, positive relationship with the consultee, in which the consultant begins to be an accessory ego figure to the latter. He does this by listening to the consultee's problems and "playing for time" in order to foster the relationship without significantly reducing the anxiety of the crisis situation. Generally this relationship can be established fairly quickly, and then the consultant begins to reduce the crisis by several techniques.

The first technique is that of "segmental tension reduction," in which the consultant fosters discussion focussed around the client's problem, which by implication is of concern to the consultee's own problems. "The consultee's segmented problem is not discussed explicitly and directly in regard to himself, but once removed in relation to the client" (Caplan, 1959). This "second-hand" or implied method of discussion helps to reduce tension within the consultee, which in turn reduces the tension within the client, due to the close psychological linkage between them.

A second technique in crisis reduction is "dissipation of the stereotype" (Caplan: 1954a, 1954b, 1956a, 1956b, 1959). This consists of the consultant focussing with the consultee upon certain realistic aspects of the client's behavior which make him more meaningful as a human being for the consultee. The consultant often does this through joint observation of the client, or in discussion in which he describes his personal views of the client and encourages the consultee to do so. This often aids the consultee to bridge the gap between his fantasy based, stereotyped perceptions and reality based, true appraisal of the client and the crisis situation which, in turn, reduces problems between consultee and client.

A major point to be kept in mind in any of these situations is the fact that the client may actually be seriously disturbed and in need of more intensive diagnosis and treatment. The consultant uses the consultation relationship to gain as many cues as possible concerning the client's behavior. He also may review recorded data as well as observe the client in conjunction with the consultee so as to gain some diagnostic clues. If the consultant feels that the client is seriously disturbed, this does not necessarily infer that the consultee does not need crisis consultation. The consultant may still employ these methods to assist the consultee in his relationship with the client, but at the same time he may shift his role to one of action-help so as to assist the consultee in making a referral of the client to an agency.

It should also be noted that the avoidance of psychotherapy is one of the cornerstones of the consultation method. There is an awareness of the personal implications of the consultee's difficulties in his work, but the latter, and not his intrapsychic conflicts, are the focus of the consultation process (Caplan: 1956a, 1959). If the consultant allows himself to get involved in the consultee's intrapsychic problems he is falling into the trap which has been set for him; namely, that the consultee was not fully certain that he is to play the role of a patient, but now he has proved that he is to do so. The consultant must do certain things which border upon psychotherapy at times, but are not necessarily the same.

The consultant develops a relationship with the consultee which is also seen in psychotherapy, although on a much slower and more intensive basis in the latter case. In psychotherapy, the therapist asks for intrapsychic material to be provided by the patient, while in consultation the consultee is actually kept from giving such material about himself. The focus is upon the client and his problems, and only through implication do some of the consultee's difficulties become involved (Caplan, 1955). In fact, the consultant bolsters the consultee's defenses by support and reassurance of his adequacy, while in psychotherapy the therapist may weaken or tear down neurotic defenses before rebuilding new ones. Another difference is the fact that the consultant does not interpret in a direct fashion in relation to the consultee's behavior, while in case insight consultation he may give some interpretation of the client's behavior which can assist the consultee in helping the client. This, too, differs quite radically from psychotherapy. Finally, in most consultation relationships there is an attempt to handle a segmental problem that faces a consultee, and once resolved, the consultant withdraws. In certain types of psychotherapy the therapist may also handle specific problems, but generally, therapy is concerned with a range of problems within the patient. It can be agreed, on the other hand, that corrective emotional experiences take place both in psychotherapy and case consultation, which may be the most important basis of comparison between the two methods. Nevertheless, one would be hard put to compare the depth and intensity of this experience in therapy with the rather narrow, segmented emotional experience that occurs through the consultation method.

A Mental Health Consultation Case

This case is a consultee crisis type. It illustrates various problems which are encountered in the initiation and follow-up of a mental health consultation relationship with a teacher. It also demonstrates how a consultant "plays for time," becomes an important accessory ego figure for the consultee (teacher), helps to reduce her tensions, and aids her in accepting a more realistic view of a client (child). This is accomplished through a reduction in the consultee's own anxieties by means of implied discussion concerning the client and his family, but which relates indirectly to the consultee's relationship with the client. The case also shows how the consultant assists in changing the consultee's stereotype about the client from one of a "bad, disturbing child," to one of a child who can be greatly aided through the consultee's understanding and support. Once these changes have been effected, the consultee is able to react toward the client in a more positive fashion, and the child's disturbing behavior is reduced.

The mental health consultant, a clinical psychologist, had been visiting a school system on a biweekly basis as part of the local mental health center's program of prevention at the elementary school level. He worked through the school social worker, who acted as a referral agent for teachers who sought his help regarding "problem children." His role appeared to have become well accepted by the school system, and he was provided a substitute teacher who accompanied him and took over the classes while a teacher was being seen in consultation.

The school social worker was most cooperative and eager to learn about mental health consultation and how he could relate his role to the consultant's. He readily complied with the consultant's approach to mental health problems and soon was making some of the best referrals from teachers in his school buildings. Although the mental health consultant received his referrals through the school social worker rather than direct self-referrals from the teachers, this did not appear to produce a barrier in his relationship with the teachers. The school social worker had clarified his role with the teachers to such a degree that he was highly accepted and often used, and this in turn aided the consultant in developing his own role.

Initial Consultation Interview

In mid-October the mental health consultant learned through the school social worker that a teacher wanted to see him about a child in her third grade class. The school social worker mentioned the fact that the child, Peter, had been doing poorly in his school work and seemed to be somewhat immature. He also said that the teacher was quite upset about the particular situation. That same day the mental health consultant visited the school and met with the teacher.

The teacher was an individual of about 50 years of age who can be best described as efficient and conscientious. The consultant realized when she (the teacher) first entered the consultation room that he had had previous contact with her. She had served as a substitute teacher for a short period of time during the previous school year and had accompanied him and taken over classes on several days when he had made consultation visits. During those visits, while riding in the mental health consultant's car, the teacher had described in great detail how she had been married for a long time but had never had any children of her own. She said that she was interested in doing volunteer or church work, and keeping up her house and garden, and it was for these reasons that she had only been teaching part-time for many years. She had decided to teach full-time this year because of the shortage of qualified teachers.

During the previous consultation visits when he had become acquainted with the teacher, the consultant was struck by the great sense of duty which this teacher felt she had "to help children to learn." When she described her role as a teacher she would often say, "Every child can learn if we really work at it with him." She also had strong feelings about the role that the home and church played in the life of the child. She decried the fact that modern parents often did not provide home life which was intellectually stimulating to the child. She mentioned how she would often take children who lived on her street into her home and talk with them about school or reading. The teacher seemed to be particularly upset about the fact that children would not learn and "wasted" their time. It seemed most difficult for her to accept this since she herself was so efficient and careful in the way that she used her time. The first description of the child by the teacher was the fact that he did "nothing at all." She went on to say that he "is emotionally disturbed, wastes his time, and has very few good work habits." She said that Peter was 7 years, 10 months old and in the third grade. He had difficulty in previous grades, but psychological tests had shown that his intellectual ability was above average and he should be able to accomplish adequately.

At this point the teacher described how she had become so upset by Peter's behavior that during the previous week she had called his father and asked that he come to her home for an interview. She admitted that the father had been angry at her request but that he had come. She described the father as very upset and disappointed with Peter's poor school work, and said that he forced the child to do homework nightly. The school social worker had also spoken to the father during a PTA meeting and had recommended that the father be somewhat closer to his son by doing things together, but the father seemed to have interpreted his remarks as forcing the child to do more homework. It seemed to the teacher that the father was being too harsh, and she told him quite directly not to force the child. On the other hand, she also told the father that she was worried about the child's poor school work and that she felt that this was related to the child's problems at home. The consultant was rather surprised that the teacher had taken matters into her own hands to such an extent. He asked the teacher how the father had reacted to her remarks. She admitted that he had been very angry at her, but as the interview continued he seemed to quiet down and said he would look into this problem.

The teacher went on to describe the home situation. She said that the mother was a very nervous person and the father was very hostile and negative. It was mentioned that Peter's younger brother was ill and had to have a number of operations. The teacher turned to the consultant and said that she thought it was a poor home. She asked what the consultant could do for her in this situation. The consultant felt that the teacher was looking for specific, concrete methods of handling the child and family, just as she might rigidly structure her teaching methods. He attempted to focus upon the child's problems and play for time. He asked about some of Peter's behavior in the classroom. The teacher described how he was very immature and aggressive. He would strike other children. When he had to sit alone as punishment he would cry and say that he wanted someone to sit with him. She also pointed out how she felt his reading and other work might improve if she would spend more time alone with him. At this point she said that she "baited" him to produce for her by offering various prizes, but this was often to no avail. He was a somewhat hostile and outspoken child, which also seemed to upset the teacher.

She turned to the consultant and again asked quite directly what he could suggest for her to do. The consultant attempted to relate the child's school behavior to home by suggesting that perhaps the schoolroom might be the place where the child could get the emotional support and warmth he seemed to need, and which he perhaps was not getting at home. The teacher did not seem to accept this point of view and instead went into a long discussion of the child's poor schoolwork.

For the third time she asked if the consultant could make any recommendations concerning ways in which he could be helped to learn. She said that another parent-teachers' conference was coming up, and she wondered what she might say to the parents. The consultant and teacher discussed the conference, and the consultant suggested that if the teacher could talk to the parents about the important interrelationship between the home and the school, perhaps they could understand why Peter was adjusting poorly in school.

The consultant hoped that this latter remark and suggestion might be accepted by the teacher, since he felt that it was the emotional climate in the school, as determined largely by the teacher's attitude and perception toward Peter, which would be very important in his future behavior. However, he could not make this a direct statement to the teacher at this time since it was related to her own segmental tensions.

At this point the teacher got up to leave, saying that she still wished that the consultant had given her more specific suggestions, but that she would talk to the parents. She hoped that the child would not become a severe learning and behavior problem some day because the school had not been able to handle him. She seemed to be very unhappy.

Although the consultant took great pains to interpret his role to the teacher at the start, nevertheless, he could see that she wished to bypass these introductory remarks and go immediately into the case in point and receive a ready-made solution. Thus, when the consultant suggested that he would discuss the problem with her instead of completely taking it over, or when he discussed matters of an emotional nature instead of giving direct suggestions for helping this boy to learn, he did not fully alleviate her anxiety. He left her feeling somewhat frustrated and angry toward him. The consultant had played for time to learn more about the situation, and had tried to develop a relationship with the teacher. He had tried to imply that the client's relationship with his family was important in his adjustment, at the same time suggesting by indirection that the consultee's relationship with the client was of major importance. The consultant was somewhat uncertain that the initial interview had been satisfactory, and he wondered if the teacher would return when she did not wish to make a definite appointment for his next visit to the school.

Later Consultation Contacts

At the next visit to the school, two weeks later, the teacher had informed the school social worker that she still did not wish to see the consultant. She said that she felt that she could handle the situation now without his aid. The school social worker mentioned this to the consultant and also remarked that the teacher seemed to be antagonistic toward any help from the consultant.

Two weeks later the consultant came to the school again, and was met in the corridor by the school social worker, who told the mental health consultant that the teacher wished to see him. He also said that the teacher was becoming so anxious over the situation that she was talking around town about the child and his family. This was very disturbing to the school social worker and the school principal, and they wondered if the consultant could handle this with her. After some discussion it was agreed that this should be handled by the school and not the consultant. The consultant felt he must have the teacher's confidence and he could not enter into an administrative relationship with her.

The consultant also felt somewhat disturbed about this discussion with the school social worker, since he wondered whether he was free to discuss the teacher and her problems with school personnel. He shifted the discussion to Peter's problems rather than the teacher's reaction to him and his family. When he did this, the social worker suddenly remarked that Peter was becoming more aggressive and was showing sexual interest by lifting little girls' skirts. This had been quite disconcerting to the teacher, and she was interested in seeing the consultant about the child. Before departing, the school social worker said that he would speak to the teacher about keeping this case more confidential.

When the consultant saw the teacher, she immediately mentioned that Peter's mother was pregnant again when she had seen her at the PTA meeting. She said that the mother was very disturbed, and while talking with her had cried continuously, telling the teacher how she did not get along well with her husband. She suddenly broke down and cried further, saying that the doctor had told her she was going to die with this pregnancy. There was some vague discussion of a relationship between pregnancy and cysts which might result in her death. The teacher was extremely upset by this whole matter and did not know what to do for the mother. The consultant suggested that the mother might be over-reacting to her pregnancy, particularly if she were the nervous person that the teacher described. This idea seemed to relieve the teacher somewhat. The consultant asked about Peter and the family, focusing upon the child again. The teacher remarked that the parents had taken her advice about Peter's needing help and had taken him to a private child guidance clinic for one visit. She did not know the results of this visit.

At this point the teacher remarked about Peter's aggressive behavior. When she asked the child why he was naughty, Peter had said that "bad voices" made him do those things. She seemed to be very fearful of this remark. The consultant interpreted this remark to the teacher, suggesting that perhaps the child was wrestling with strong and upsetting feelings and this was his way of expressing them more openly. The reason for this interpretation was that the consultant felt the teacher was quite disturbed by the boy's remark. The consultant was trying to alleviate some of her anxieties about her own impulses, which appeared to be closely related to her identification with Peter. The teacher said that she had told Peter to ask God for help, and she was very fearful of what his remark had meant in terms of a "mental disturbance." She did not remark about Peter's sexualized behavior.

The teacher suddenly changed the subject and started describing the family again, telling how they were rather strange people who did not get along well with others and were even rejected by their neighbors. After some discussion, the consultant suggested that this was useful background information in terms of the child's basic school problem. The consultant, focussing on the child again, went on to remark that, actually, the fact that Peter was able to express these inner feelings to the teacher seemed to be a positive and important factor in helping him. He said that if Peter really did feel he could tell the teacher about these bad thoughts it was because he basically had a good relationship with her. The teacher seemed to react positively at this remark. She said that perhaps she had been wrong and had misunderstood and mishandled him at times. She seemed to be looking to the consultant for his approbation. The consultant suggested that all teachers feel conscientious in their work, and want to help children such as Peter. He went on to say that it was really not a poor reflection on teachers if things did not move smoothly, since this was really a difficult task. The teacher seemed to be more relieved of her personal anxiety and began to remark how unhappy a child he really was. The consultant agreed and said that sometimes children were seen as bad when actually they were disturbed, unhappy persons. Again the teacher agreed, and made several remarks about the boy's basic feelings of unhappiness. At the end of the interview she thanked the consultant profusely, saying that she now felt much less anxious about the situation. She also said that she would like to talk further with him about Peter. The consultant made an appointment with her to return in two weeks.

At this stage in the development of the consultation case, the consultant listened to the consultee discuss the client's family and express her thoughts and anxieties about their neurotic difficulties. By implication, she also expressed her own feelings of inner unhappines and disturbance, which the consultant noted. This rose to a peak when the consultee discussed the client's "bad voices," but the consultant did not deal directly in a therapeutic fashion with the consultee's problems. Instead, he bolstered her defenses indirectly through supportive statements, to which she responded with more understanding remarks about the client and his problems. This suggested that her own tension may have begun to wane.

On his next visit, the teacher immediately launched into a description of Peter during the past two weeks. She said he had been out sick for a while, and when he came back he had begun to throw some temper tantrums. She had been concerned with this behavior and had tried to talk with him, and he had stopped within a few days. She also mentioned that he had been writing love notes to a little girl. This, too, had been disturbing to her, and she had threatened Peter, saying that she would punish him severely or tell his parents if he continued to do this. It appeared to the consultant that she did not attempt to gain any understanding of Peter's motives in such behavior. The consultant wondered why Peter might be doing these things. The teacher did not wish to discuss his motivation at this time, but instead she began to talk about the child's mother again. She showed a letter to the consultant from his mother to her. The letter had some rather strange wording in it, upon which the consultant remarked. There was some discussion of the mother as a very anxious and somewhat peculiar individual.

The consultant felt that once again the teacher was becoming too closely identified with the family and with the pregnant mother in particular. He tried to shift the focus back to the child by remarking that perhaps the mother was a very disturbed person, but at the same time the school was faced with the problem of Peter and his need for support from that source. The teacher seemed to be stimulated by the word "support" and mentioned how he did much better in reading when he sat near her. The consultant picked up this point and re-emphasized the child's need for a close relationship. The teacher felt that perhaps she could try to work more closely with him. As she said this, the consultant was not certain how close she could let herself get to this frightened child and his family, although he felt the teacher had made some headway in this regard. The rest of the consultation session was spent in discussing how a disturbed child may sometimes look for other sources of comfort and nurturance if he cannot receive this from his family. The consultant again re-emphasized the fact that the teacher seemed to be a very important individual in the boy's life, and she seemed to be supported by this remark. During this interview, the consultant continued to reduce the teacher's anxiety, and through his increasing relationship he made some more specific remarks about the teacher's role in relation to the child. The teacher began to discuss upsetting sexual material, but her wish to drop the subject was respected by the consultant.

The next visit was made a month later.

The teacher immediately remarked on how improved Peter was since the consultant's last visit. He had stopped writing notes, and he seemed to be settling down to his work. She told of an incident where children were making placemats as Christmas gifts. Most were making them for their mothers, but in Peter's case he remarked it was for the little girl to whom he had been writing love notes. This was discussed, and the consultant suggested that perhaps Peter needed love to such a degree that he looked to other female figures, such as little girls. At the same time, he suggested that mother's pregnancy might also be motivating him toward sexual interests. The teacher seemed ready to listen, and she accepted these remarks, saying that she had never really understood the reasons for this behavior until now. She also spoke about the dress-lifting episodes for the first time, and said that maybe he was doing this because of his mother's pregnancy, too.

She continued on with some more positive statements with regard to Peter. She remarked how he had "guarded" a small Christmas tree in the school room when she had suggested that he be the person to watch over it. He seemed to like this role very much, and the teacher was pleased that he had done a good job. It seemed that the teacher was much less anxious and tense about the case at this point. She discussed it from a much more general viewpoint, saying that now she understood the family background and how unhappy the child had been. She wanted to help him and her goal was to at least start where he was and see where she could get with him. The consultant pointed out that things seemed to be moving along fairly well at this point. The teacher smiled happily and asked if she was doing a good job, saying that she had not taught young children for a long time before this past year and was not sure if she could really work well with children of this age. The consultant remarked that he could see she was trying to create a good emotional atmosphere and relationship with Peter in order that he could carry on his learning. At this point the teacher made many positive remarks about the parents, saying that they seemed much more cooperative now. The meeting ended with some general discussion about mental health consultation, with the teacher saying that she felt much better about the whole situation. When she left, she said she did not know if she would have to see the consultant any more.

This was the last time that the consultant saw the teacher in formal mental health consultation. However, on an informal basis, he would sometimes see her in the school lunchroom. She remarked through the remainder of the school year that the boy seemed to be adjusting much better and was keeping up with his school work. In the spring, Peter's mother had a baby and was in good health. It was felt by the consultant that something had been accomplished in helping the teacher clarify her interpersonal relationship with this boy as well as her view of his family. In this final interview, the consultee appeared to be able to accept and deal with the client in a much more positive fashion, and she is more understanding of his needs. At the end, she indirectly stated that she had become more understanding and cooperative as she discussed the client's family. The stereotype was reduced through discussion, and all were accepted more warmly.

Evaluation of Consultation

In four contacts, this very conscientious, anxious teacher seemed to have been able to work through some of her own feelings with regard to a child and family who presented many problems of a disturbing nature to her. There were several features in this case which were of particular interest. The teacher was operating at a high level of anxiety and concern, and she directed her energy into maladaptive channels. She talked very openly and freely concerning the problem outside the school; she attempted to force the parents to do her bidding, and she reacted in a negative fashion to the child. In addition, this teacher was extremely critical and upset with the client's family since it did not fit into her rigidly developed set of social values. There were strong suggestions that the child's hostility, emotional disturbance, verbalization of "bad voices," and sexualized behavior were quite disconcerting to the teacher. She reacted by punishing him or by subtly rejecting him, although her conscientiousness and anxiety caused her to be constantly involved in his problems in an attempt to find solutions.

The mental health consultant appeared to have done a number of things which seemed to have helped in this situation:

He focused upon the child as the primary area of concern for the teacher in an attempt to divert her energy from acting out with the parents or outside the school.

He attempted to interpret the child's basic needs to the teacher in such a way that she could begin to accept him, not through the stereotype of a "bad boy," but through more objective perception of a disturbed, unhappy child who needed a stable adult relationship.

He tried to help the consultee feel that if she actually did come close to this child she would not be contaminated with his disturbing influence, but instead, that she would be playing an important role in helping him.

The consultant helped the teacher feel more secure in her role as a teacher of young children, when she had not taught them for a long time, by implying that being conscientious and working hard at her job was not necessarily a poor trait but very useful to children, especially if this provided them an emotional atmosphere in which they could learn.

On the other hand, there were things that the consultant did not do:

He did not become involved in the school hierarchy and have to play the role of "punishing" the teacher for her breaking of case confidentiality.

He did not go into the meaning of "bad voices" and the deep anxieties this engendered in the teacher. Instead, he only highlighted certain positive aspects of this behavior for the teacher in an attempt to help her reduce her own anxiety. At the same time, the fact that she had used religion as a method of control might have been underscored somewhat by the consultant as a socially acceptable defense mechanism which would have supported the teacher still more.

He did not become involved in the "traps" set by the teacher when she asked for specific methods of handling the child and his family. This might have reduced the crisis situation by the usual supportive or manipulative methods without any more useful long-term resultant of a preventive nature. Instead, he constantly focused on the child's problems and helped the teacher work through her feelings concerning her relationship to the child.

Finally, it should be stressed that work in this area of mental health consultation demands psychological skill and the ability to control one's own anxiety. The flow of information must be evaluated with regard to its degree of reality distortion. The social system of the school building, school system, and larger community must be kept in mind as a broad frame of reference. The consultant must be continually aware of new cues, changes in behavior, or other psychological manifestations. Last, but not least, he must know how to apply his consultative techniques to their best advantages so as to remain a key member of the mental health team in the development of preventive approaches to emotional disturbances in school children.

REFERENCES

BAKER, EDITH, & SIEGEL, DORIS. *Medical social services for children.* Washington: Children's Bureau, 1953.

CAPLAN, G. Principles of mental health consultation. Unpublished manuscript. Harvard U. Sch. Pub. Health, Boston, Mass., 1954. (a)

CAPLAN, G. Dynamics of mental health consultation. Unpublished manuscript. Harvard U. Sch. Pub. Health, Boston, Mass., 1954. (b)

CAPLAN, G. The role of the social worker in preventive psychiatry. *Med. Soc. Wk.*, 1955, **4**, 144–159.

CAPLAN, G. *Mental health aspects of social work in public health.* Washington: Children's Bureau, 1956. (a)

CAPLAN, G. Mental health consultation in schools. In *The elements of a community mental health program.* New York: Milbank Memorial Fund, 1956. Pp. 77–85. (b)

CAPLAN, G. *Concepts of mental health and consultation.* Washington: Children's Bureau, 1959.

HALLOCK, A. K., & VAUGHAN, W. T. Community organization—A dynamic component of community mental health practice. *Amer. J. Orthopsychiat.*, 1956, **26**, 691–706.

PROSSER, D. The summary of mental health consultations which seem to emerge out of the work of the four subgroups. In *Proceedings of the ninth annual conference of chief social workers from state mental health programs.* Bethesda: National Institute of Mental Health, 1957.

TOWLE, CHARLOTTE. *The consultation process.* Unpublished manuscript. Univer. of Chicago, 1950.

VAUGHAN, W. T. Mental health for school children. *Children*, 1955, **2**, 203–207.

(Received January 2, 1959)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

THE RELATIVE CONTRIBUTION OF FOUR KINDS OF DATA TO ACCURACY IN PERSONALITY ASSESSMENT[1]

LLOYD K. SINES[2]

Fergus Falls State Hospital, Minnesota

Does the information available to the clinical psychologist through his tests and other resources, as in the psychodiagnostic process, provide him with a significant amount of understanding of the patients about whom he is asked to make judgments, descriptive and diagnostic statements, and for whom he is asked to make decisions? This question implies others: (*a*) if significant insights into patients are possible via the information available to the psychologist, which kinds of data contribute most to an adequate level of understanding, (*b*) what is the relationship between the amount of data available to the clinician and the degree of insight he achieves, and (*c*) is it possible to identify most efficient batteries of tests and/or other data. More precisely, we see that it is not only the data themselves in which we are interested but the data as they are used by the psychologist. As Kelly (1954) has noted, the introduction of the human element into the assessment process means that the techniques of assessment (tests, data) have validity which is not independent of the assessor. The problem is, then, one of evaluation of interaction between data and user.

Hypotheses

1. Clinicians are able, on the basis of psychological test, history, and interview data, to describe psychiatric patients more accurately than does a stereotyped personality description obtained by reference to the prevailing conceptions of the typical or "average" psychiatric patient.

2. (*a*) Psychological test, history, and interview data are not of equal value in terms of their contribution to diagnostic and descriptive accuracy.

(*b*) Of the psychological tests used, that which is designed to tap the inner, unconscious, and genotypic facets of personality yields a greater degree of insight into patients, and a higher level of accuracy of judgments.

3. There is a positive relationship between the amount of data available about a patient and the accuracy with which judgments, inferences, and diagnostic statements can be made about him. The information conveyed about a patient by each kind of data is additive, each succeeding datum increasing the accuracy of judgments.

4. Although there are differences among the kinds of data with respect to their usefulness or "validity," clinicians differ with respect to their success with the available techniques or data.

5. Because psychodiagnostic techniques (especially psychological tests) reveal the covert, often unconscious apsects of personality, the judgments and inferences made about patients' personality characteristics by psychodiagnosticians will agree more closely with psychotherapists' perceptions of those patients

---

[1] This paper is based upon the doctoral dissertation of the author at the University of Minnesota. He wishes to thank William Schofield for his invaluable assistance and encouragement in planning and completing the study. Also, the services of the trainees and staff at the VA Mental Hygiene Clinic, Fort Snelling, Minnesota, were indispensable and especially appreciated.

[2] Formerly of the Veterans Administration, Ft. Snelling, Minnesota, and currently affiliated with the University of Minnesota.

after more extended therapeutic contact with them than with the therapists' impressions after only a moderate amount of such contact.

Method

Patients and instruments. Applicants[3] to a Veterans Administration mental hygiene clinic were screened during the intake process for (*a*) judged "stayability" in individual psychotherapy, and (*b*) availability for a psychological evaluation of approximately six hours in length. The psychological examination consisted of the following procedures: each patient completed a four-page Biographical Data Sheet (BDS) and the booklet form of the Minnesota Multiphasic Personality Inventory (MMPI). He was also interviewed and given a Rorschach test by the clinical psychologist who was the "diagnostician" and who performed the subsequent clinical Q sorts.

The data were made available to the clinician sequentially, and he performed a Q sort of 97 items describing phenotypic and genotypic personality and behavioral characteristics after considering each of the four major groups or kinds of data. The biographical data were uniformly made available to the clinician as the first group of data, and the initial Q sort was performed on that basis alone. Six sequences of data were possible with the remaining three kinds of data, and in order to exhaust the number of sequences each of the (five) diagnosticians examined six patients, each patient being studied by a different sequence of data. The sequence followed in a given clinician's first case was assigned by random selection, and each subsequent sequence by random selection from among the remaining (unused) sequences. It was understood that the four Q sorts in each instance would not represent interpretations of the specific data which immediately preceded them, but a summation of all the data which has been made available to the clinician up to that point in the diagnostic process. Thus, the fourth (final) clinical sort represented the clinician's integration of all the data included in the battery, rather than his interpretations of the data which were made available after the third Q sort had been performed. An 11-category quasi-normal forced distribution was used in performing the Q sorts.

After the patient had been seen for a total of 10 interviews,[4] his therapist[5] performed a (criterion) Q sort of the 97 items used by the diagnostician in describing the patient. Comparisons between the diagnosticians' four clinical sorts and the therapists' 10th-hour Q sort descriptions of the patients constitute the major data of the study. In addition, for those patients who remained in therapy for 15 or more interviews, an additional Q sort description of them by their therapist was obtained at the time the study was terminated. The range of therapy interviews in this group was from 15 to 49. Comparisons between Q sort descriptions obtained at that time and the 10th-hour Q sorts as well as between the former and the diagnosticians' four clinical sorts were made. As a further methodological consideration, in 10 of the 30 cases a re-sort from the therapist was obtained between the 10th and 11th therapy interviews or after the terminal Q sort was obtained in those cases remaining in therapy beyond 15 interviews. The correlations between the 10th-hour Q sorts and the re-sort, and between the terminal Q sort and its re-sort are considered a measure of reliability.

Five clinical psychologist trainees comprised the group of diagnosticians (hereafter referred to as clinicians). They were at different levels of training (within the limits of the doctoral training program), but had all been employed by the VA as trainees for at least one year.

Selection of the Q sort item pool. The item pool finally selected for use in the present study was obtained from a larger pool[6] consisting of 295 items gathered from a variety of sources, including the studies by Halbower (1955), Block and Bailey (1953), and others. Further, each of the Murray (1938) needs was included as a separate item, as were the psychoanalytic defense mechanisms. Various of the diagnostic categories found in the *Diagnostic and Statistical Manual* (American Psychiatric Association, 1952) were represented by items, as were the personality and behavioral characteristics associated with the more common MMPI code types described by Hathaway and Meehl (1951).

Each of 11 psychology trainees performed a Q sort describing their conception of the typical MHC patient (referred to psychology for testing). These sorts were summed to form a composite from which was derived a Mean Average Patient (MAP) description. The MAP description was given to each of the five clinicians in the study who dichotomized the distribution at the point where he felt that the items became, by and large, *false* with respect to the typical MHC patient. (This procedure was also employed by the clinicians when performing the fourth clinical sort for each patient.) The stereotype (MAP) description was compared with the criterion (therapist) Q sort for each patient. The average level of agreement between the two over the sample of 30 patients was considered the baseline (chance, or expected level) against which the clinical sorts were evaluated.

---

[3] The patients ranged in age from 24 to 47 years (median 32.5) and in education from 6 to 19 years (median 12.0). Approximately two thirds had disability ratings for a psychoneurotic condition and about one fifth for a psychotic condition. The degree of impairment ranged from 0% to 100% (median 10%).

[4] In three cases the therapist's Q sort was obtained after the eighth therapeutic interview because of the conclusion of the study.

[5] Thirteen therapists drawn from each of the three major disciplines in the clinic (7 clinical psychologists, 4 psychiatrists, and 2 social workers) were involved in the study. About one half of the patients were seen for therapy by psychologists.

[6] This pool was made available to the author by P. E. Meehl. The 97 items used in this study may be made available upon request.

The method and results of the Q sort item selection were as follows: four pairs of judges (one pair for each of the four data groups included in the diagnostic battery) rated each of the 295 items in the original pool. Each item was judged as "suitable," "of questionable suitability," or "unsuitable" to rating, on the basis of the information conveyed by the specific kind of data. The judges' ratings were arbitrarily assigned weights of 2, 1, or 0, respectively, so that each item had a possible score of 0 to 4 for each data group, and from 0 to 16 as an over-all rating. Since it was desirable that the items finally selected have relevance to each of the kinds of data utilized by the clinicians, items were selected from the larger pool on the basis of (*a*) a high total rating score, and (*b*) specific pertinence (judged suitability) to one or more of the four types of data to be used by the diagnosticians as the basis for their subsequent sortings. A total of 94 items was so selected, and two of the items were modified and expanded into five separate items, thus comprising the final 97-item pool used in the study proper. An attempt was made to equate the final pool with items judged to have particular pertinence to each specific kind of data. The extent to which an equating of the pool for items pertinent to each data type was accomplished is indicated by the data in Table 1. From Table 1 it may be seen that the Interview, MMPI, and Rorschach were roughly comparable in terms of their item representation in the Q pool, while the BDS was distinctly underrepresented by items judged to have pertinence to that source of data.

As a further measure of the representation in the item pool of items relevant to each of the data types, the subsamples of items contained in the final pool which were rated by each of the two judges as "suitable" to a specific data type may be compared with the total pool in terms of item placements (means and variances) in the stereotype (MAP) sort. The 97-item Q sort distribution utilized in the final sortings yielded a mean of 6.0 and a variance of 4.54. The means of the four subsamples of items were as follows: Interview, 5.52; MMPI, 5.36; Rorschach, 5.61; and BDS, 7.53. The variances were: 5.06, 4.28, 5.56, 5.05, respectively. An analysis of variance of these values yields an F of 3.76, which indicates statistically significant differences among the item subsample means. By t test analyses it was shown that the BDS item subsample differed from the remaining three samples in the mean placement of items in the MAP sort. It would appear then, that the 97-item Q pool was underrepresented with items inferable from the BDS, and further, that the small sample of items "suitable" to rating from the BDS which was included in the total pool could not be considered a random sample from the total item pool, but rather, contained items which tended to be placed toward the less descriptive end of the Q sort continuum in the MAP sort.

Table 1

Judges' Ratings of the 94 Q Sort Items Originally Selected from the Larger Pool in Terms of Their "Suitability" to Rating from the Specific Types of Data Utilized by the Clinicians

	Sum of the Two Judges' Ratings						
Kind of Data	4	3	2	1	0	Sum × Frequency	%
Interview	44	29	16	4	1	299	28.8
MMPI	45	22	16	5	6	283	27.3
Rorschach	43	25	10	12	4	279	26.9
BDS	15	16	20	25	17	176	17.0

Results and Discussion

Reliability of the Q sort. The reliability sorts were performed by the therapists on different days, and were not separated in time by a therapeutic interview with the patient. The number of days between initial sort and re-sort for the 10 cases ranged from 1 to 42, with a median of three days. The two sorts were correlated and yielded a range of reliability coefficients from .795 to .936 (median .858, mean .866). There was little apparent relationship between the magnitude of the reliability coefficient and the length of the time interval between sorts, perhaps because of the small dispersion of the latter. These data are interpreted as indicating satisfactory reliability of the Q sort.

Results of testing Hypothesis 1. (*a*) In 21 of the 30 cases the correlation between the final clinical sort and the therapist's criterion description for that patient exceeded numerically the correlation between the MAP description and the latter. The sign test indicates that this proportion would arise on the basis of chance alone in less than 5 out of 100 instances. (*b*) In 15 of the above 21 cases the observed differences between correlations were statistically reliable ($P < .05$). (*c*) In 9 of the 30 cases the MAP (stereotype) description exceeded the validity of the final clinical sort. Four of the nine differences were statistically reliable. (*d*) In 21 of the 30 cases the number of items agreed upon by clinician and therapist on the true–false dichotomy exceeded the comparable figure in the case of

Table 2

Validity Coefficients of Six Subgroups of Data

	Biographical Data Sheet *plus**			Complete Battery *minus**		
	Interv.	MMPI	Rorsch.	Interv.	MMPI	Rorsch.
Mean validity of the stereotype	.272	.384	.347	.343	.348	.294
Mean validity of the BDS alone	.381	.385	.411	.402	.403	.366
Validity of the data subgroup						
Range	.282 to .734	.120 to .539	.016 to .550	.016 to .627	.220 to .607	.341 to .698
Median	.610	.410	.395	.456	.453	.604
Mean	.566	.378	.368	.403	.450	.595

* $N = 10$ different cases in each subgroup, thus exhausting the total sample of 30 patients.

the MAP description/therapist comparison. Again, the sign test yields a probability of less than .05, which indicates a significant difference in favor of the clinical sorts over the MAP description. (*e*) The average percentage of the 97 items agreed upon (over the 30 cases) by clinician and therapist was 66.0, whereas the comparable figure in the case of the MAP description/therapist comparisons was 60.0. The difference between these two means is statistically reliable ($P < .05$). (*f*) For the entire sample of 30 cases the average validity coefficient of the final clinical sort was .480, while the average validity coefficient of the MAP (stereotype) description was .340. The difference between these mean values is statistically reliable ($P < .01$).

In each of the above analyses, evidence of the superiority of the clinical sorts over the stereotype description is observed. Although the final clinical sort (which represented a clinician's integration of all the available data on a patient) did not *always* result in better than chance descriptive accuracy, it did so in a majority of the cases, and in one half of the cases the greater accuracy of the clinical description was statistically reliable. Hypothesis 1 is accepted.

Results of testing Hypothesis 2. The validity coefficients of six batteries or subgroups of data are reported in Table 2. The data indicate that although the subsamples of patients seemed to be roughly comparable in terms of the degree to which their BDS contributed to descriptive accuracy, and they were not markedly dissimilar with respect to their "typicalness" (MAP validity), quite different validities accrued from the data (MMPI, Rorschach, or Interview) which were considered by the clinicians following the BDS sorts. None of the BDS *plus* MMPI or BDS *plus* Rorschach validity coefficients equaled or exceeded the median coefficient obtained in the BDS *plus* Interview group. Further, the mean validity in the latter group (.566) was significantly greater than the mean in either of the other two groups ($P < .05$). Also, only the BDS *plus* Interview mean validity was reliably greater than the average validity of the MAP description for that subgroup ($P < .01$) and, in addition, significantly greater than the average validity of the BDS alone for the 10 cases in the subgroup. Further, considerable differences were observed at the point of the third clinical sort, depending on which kind of data was omitted from the battery. In only one case from each of the Battery *minus* Interview and Battery *minus* MMPI groups did the validity coefficient equal or exceed the median coefficient in the Battery *minus* Rorschach group. Further, although the average validity coefficient in each of the three groups exceeded the average validity coefficient of the MAP description for the 10 cases in that group, in only the Battery *minus* Rorschach sample was the observed difference statistically reliable ($P < .01$). Finally, the mean Battery *minus* Rorschach validity coefficient (.595) was significantly greater than the mean validity coefficient in either of the other two groups ($P < .05$). In view of these facts, and (as may be seen from Table 2) since the average validity coefficient of the BDS *plus* Rorschach did not exceed the average validity coefficient of the BDS *plus* MMPI, Hypothesis 2_b, which implies the greater validity of the Rorschach relative to the MMPI, is not confirmed.

A further method of elucidating the relative power or usefulness of the data used in the diagnostic process is reflected in Table 3. The frequency with which each of the kinds of data immediately preceded the Q sort which yielded the highest validity coefficient of the four obtained during the diagnostic process was tabulated. In addition to the four kinds of clinical information, the MAP description is included in the tabulations. Finally, the ranks of one through five are given (inverse) weights and, when multiplied by the observed frequencies and summed, the data are ranked in terms of their contribution to the maximum validity obtained during the diagnostic process.

Several important trends may be observed in the data reported in Table 3. First, the frequency with which the Interview yielded the greatest agreement with the therapist (40% of the cases) is striking. In addition to a possible explanation in terms of the similarity between this data type and the criterion (the fact that diagnosticians' interview impressions were correlated with therapists' interview impressions), the possibility is suggested that the diagnostic interview was in itself a quite useful technique which contributed at least as much to the diagnostic process as the best available "test" in the conventional sense. A second important fact reflected in Table 3 is the distribution of the MAP validities. It would appear that the stereotype description tended either to fit the patient very well (better than the tests) or to miss him almost completely, so that any specific test or history data about him afforded a more accurate description than did the stereotype. In 27% of the cases no single test or combination of data led to a more accurate description of the patient than was afforded by the stereotype. That the latter did not obtain by virtue of the tests' failure to describe the patients well, but rather, was due to the absolute accuracy of the stereotype description in many instances was attested to as follows. In the eight cases where the latter yielded the highest validity coefficient, the correlations ranged from .368 to .705 with a mean of .550, which is just about as good as even the best test batteries did in a given series. There appeared then, to be a subpopulation of patients for whom the MAP description was quite accurate, and for whom little further data were needed in order that they be described as adequately as the psycholo-

Table 3

The Frequency with which Each of the Kinds of Data, and the Stereotype, Immediately Preceded the Highest to Lowest Validity Coefficients in Each of the 30 Diagnostic Cases

Magnitude of Validity Coefficient	Weight	Interv.	MMPI	Rorsch.	BDS	MAP
Highest	4	12.0	5.0	1.0	4.0	8.0
2nd	3	8.0	7.0	8.0	5.0	2.0
3rd	2	5.5	8.0	13.0	3.5	0.0
4th	1	1.5	7.0	6.0	10.5	5.0
Lowest	0	3.0	3.0	2.0	7.0	15.0
Sum of Weight × Frequency		84.5	64.0	60.0	48.5	43.0

Note.—$\chi^2 = 58.39$; $df = 16$; $P < .01$.

gist was able to do on the basis of an extended workup. On the other hand, there was a larger subgroup which was quite unlike the MAP description, and for whom additional data were needed in order that they be accurately described or diagnosed.

The data presented in this section, plus several of the comparisons discussed in the previous section, may be summarized by means of a ranking of the batteries reviewed above in terms of their average validity coefficients (Table 4). The average validity of the MAP description may be construed as the baseline or chance level of accuracy.

From the preceding, it may be seen that the kinds of data used by the clinician differed with respect to their contribution to diagnostic accuracy. By way of summarizing the relationships among the kinds of data used it would appear that (*a*) the superior batteries (in terms of agreement with the therapist) were those which included the Interview, (*b*) omitting the Rorschach test from the battery was associated with higher validity of personality descriptions than when it was included, and (*c*) heavy reliance on the Rorschach was associated with no better than chance accuracy of personality descriptions. Hypothesis 2_a, which postulates differences among the data in their contribution to accuracy in personality description, is accepted.

Results of testing Hypothesis 3. Hypothesis 3 implies that as data are added during the course of the diagnostic process the accuracy of judgments about the individuals studied increases. Further, it was hypothesized that each kind of data used in the study increases the validity of the subsequent Q sort over that of the previous Q sorts as it is added to the battery.

Table 4

Average Validity Coefficients of the Several "Batteries" of Data

Battery (Basis of Q Sort)	N	Average Validity Coefficient
Battery *minus* Rorschach	10	.595
BDS *plus* Interview	10	.566
Complete Battery	30	.480
Battery *minus* MMPI	10	.450
Battery *minus* Interview	10	.403
BDS alone	30	.396
BDS *plus* MMPI	10	.378
BDS *plus* Rorschach	10	.368
Stereotype (MAP)	30	.340

Table 5

The Relationship between the Ordinal Position of the Q Sorts and Their Average Validity Coefficients

(N = 30)

Q Sort	Mean Validity Coefficient
Individual clinician's "average patient" sort	.289
MAP description	.340
Clinical sorts	
1st	.396
2nd	.446
3rd	.477
4th	.480

From Table 5 there does appear, over the entire sample, to be a consistent positive relationship between the amount of data concerning patients that clinicians had on which to base their judgments and the accuracy of such judgments. However, in spite of the consistent trend toward increments in validity as a function of the amount of data available, the absolute improvement from initial (BDS) to final clinical sort was quite small. While, on the average, the BDS sorts accounted for about 15% of the criterion variance, the sorts on the basis of the entire battery accounted for less than 25% of that variance—an absolute gain of less than 10% fewer errors in judgments (accuracy of description).

Although the actual gains in validity with additions of data were minimal, there is very good evidence that clinicians' perceptions of patients rapidly crystalize as a function of the amount of data (Table 6).

Table 6 shows clearly that the clinicians changed their impresions of a patient less and less with further increments to the available data. While they may have changed from their "average patient" conception on the basis of a patient's BDS, they were unwilling to greatly alter their statements after they had considered two or three kinds of data about a patient. In view of the rapid crystalization of clinicians' perception of patients, and only slight improvement in validity as a function of increments in the amount of data available, Dailey's (1952) finding, that clinicians in this type of enterprise begin to interpret their early inferences as *facts* as they are required to make further judgments, appears altogether plausible.

By way of further analysis, the number of times the addition of the MMPI, Interview, and Rorschach, respectively increased or decreased the validity of the clinicians' judgments may be interpreted as evidence of their contribution to the diagnostic battery. Considering the absolute numerical value of the validity coefficients as the basis of comparison, it may be seen that a given Q sort validity should either exceed or be less than the validity of the preceding Q sort in an equal number of cases if there were no differential value of the specific data class in the battery. The chance frequency with which each data class should yield a higher or lower validity coefficient than the immediately preceding sort is 15. Tabulation of the actual frequencies shows that the Rorschach Q sort yielded a (numerically) higher validity coefficient than the Q sort immediately preceding it in 10 cases, the MMPI in 13, and the Interview in 25. According to the sign test of significance, only the last value is reliably different from the expected value of 15. These data indicate that whether or not the data class added to the validity of an existing battery depended on the nature of the data. In this analysis, only in the case of the Interview did an increase in validity over that obtained on the basis of previously considered data accrue in a significant number of cases.[7]

Table 6

Average Intercorrelation of Successive Q Sorts by the Diagnosticians

Q Sort	Mean Intercorrelation
Clinician's "average patient" sort vs. his 1st clinical sort	.459
1st clinical sort vs. 2nd clinical sort	.698
2nd clinical sort vs. 3rd clinical sort	.805
3rd clinical sort vs. 4th clinical sort	.850*

* Approximates the median reliability coefficient of the Q sort.

Table 7

Validity Coefficients of the Interview when it was the Second, Third, and Fourth Data Group Considered in the Diagnostic Process

(N = 10 in each sample)

Ordinal Position of Interview	Average Validity Coefficient of Sort	Average Validity Coefficient of Final Sort
2nd	.566	.542
3rd	.480	.477
4th	.412	.412

Finally, although the Interview tended to add to the accuracy of judgments made on the basis of test and other data, the extent to which it contributed depended upon how much data about the patient had already been made available to the clinician. That is, if the Interview was held early in the diagnostic process a much greater contribution to the validity of the clinician's judgments was made than if the Interview was held later, or was the last procedure employed. Since there is evidence indicating that clinicians tended to crystalize their thinking about patients early in the diagnostic process regardless of the specific data on which their judgments were made, it appeared that if the Interview was a part of the early body of data on which closure was based, a relatively high level of validity obtained, not only at that point, but also at the completion of the diagnostic process. However, though the Interview was a powerful diagnostic technique, it could not overcome the false leads obtained early in the diagnostic process on the basis of the less valid measures or techniques. Table 7 contains the data on which these conclusions are based.

In summary, the relationship between the amount of data available to clinicians and the accuracy of their judgments appeared to be complex rather than as simply stated in the hypothesis. When all cases and all data groups were considered, there appeared to be a consistent trend toward increased accuracy (validity) as further increments to the data were made. However, the above relationship is seen to depend largely on the nature of the data which were added to the battery. In the case of the Rorschach, decrements in validity were as frequently observed as were increments when that test was added to the battery, while the Interview almost invariably added to the accuracy of the diagnostic judgments, the degree of the latter being related to the point in the diagnostic process at which it occurred.

---

[7] It should be noted that sequence or ordinal position in the series was equated for all groups, since the values were computed on the basis of all 30 cases. Each data type, then, appeared as the second, third, and fourth group an equal number of times.

Results of testing Hypothesis 4. Hypothesis 4 states that individual differences would be observed among the clinicians in terms of their success with the kinds of data utilized. The analysis in this connection is a rather straightforward one, namely the tabulation of the frequency with which each clinician's highest validity coefficient in each of his six diagnostic cases followed the introduction of each of the four data types. Table 8 contains the pertinent data.

First, it should be noted that with the exception of the BDS, which was routinely analyzed first and without benefit of other test data being available, sequence or ordinal position in the diagnostic process is not a significant factor in the distribution of validity scores, since the figures were computed on the basis of the entire sample of 30 cases, which indicates that each of the MMPI, Rorschach, and Interview was second, third, and fourth an equal number of times (10) in the sequence.

Table 8

The Frequency with which Each of the Data Groups Produced the Highest Validity Coefficient in the Six Cases Studied by Each Clinician

Clinician	VA Training Classification (year)	Highest Validity Coefficient			
		BDS	MMPI	Rorschach	Interview
A	IV	1	3	2	0
B	III	2	0	0	4
C	III	2	2	1	1
D	III	0	0	0	6
E	II	2	1	0	3
Total		7	6	3	14

The cell frequencies in Table 8 may be evaluated for significance by means of a nonparametric test. Since, for each clinician, there were six trials (patients) and four datatypes, the mean expected frequency per cell is 1.5. Further, the probability that a specific data group (test) would yield the highest validity coefficient in a given case is one in four (.25), of the same data group being highest in any two cases (.0625) (15) = .9375, in any three cases (.015625) (10) = .156, in any four cases (.00390625) (10) = .039, in any five cases (.0009765625) (4) = .004, and in all six cases .0002. From these facts it may be said with statistical confidence that clinicians B and D were more successful with the Interview, relative to their use of the other data, than were clinicians A and C, relative to the latter's use of the other data. Although these relationships are quite clear from a descriptive point of view, it is not possible from the data to explain the relationships, or to make inferences to over-all diagnostic (or even interview) ability. It may be seen in this connection that the fact that clinician D obtained his greatest validity on the basis of the Interview in each instance could have come about as readily from a lack of skill with the other data as from some unique interview acumen.[8]

A further matter relating to individual differences among the clinicians concerns their over-all level of accuracy in the diagnostic enterprise. An evaluation of this factor was accomplished as follows: the 30 validity coefficients obtained by the five clinicians were ranked in terms of their magnitude, and each clinician's distribution of six coefficients was obtained. An analysis of variance was applied to the five distributions of validity co-

---

[8] A further interesting possibility is suggested by the data in Table 8, namely, that a trend indicating a relationship between the amount of experience (level of training) of the clinician and his degree of success with psychological tests may exist. Though the range of experience among the clinicians was relatively small, it is interesting to note that only in the case of the most advanced trainee were the tests used to particular advantage.

Table 9

Validity Coefficients of the Five Clinicians' Q Sort Diagnostic Judgments

Clinician	VA Training Classification (year)	Validity Coefficients		
		Range	Mean	Median
A	IV	.361 to .659	.531	.533
B	III	.327 to .652	.475	.460
C	III	.293 to .705	.535	.523
D	III	.125 to .627	.458	.528
E	II	.123 to .466	.387	.429

efficient ranks, and yielded an F of 12.8 indicating a probability of less than .001 that the distributions of ranks represented random samples from a common source. A t test analysis of the individual distributions of ranks indicated that clinician E produced validity coefficients significantly lower than those of the other four clinicians. Further, only two of clinician E's six validity coefficients exceeded the lowest median coefficient obtained by any of the other clinicians. Although it would have been desirable to relate the degree of experience and the level of the clinicians' training to their relative success in the diagnostic enterprise, the range of experience (training) among the five was quite restricted. However, as indicated in Table 9, the fourth year trainee had the highest median validity coefficient, while the second year trainee had the lowest coefficients. Ideally, of course, a critical evaluation of the relative power or usefulness of various kinds of data, and a definitive statement of the relationships between amount of experience (or level of training) and diagnostic accuracy, would require study or a group of thoroughly trained, broadly experienced clinicians, rather than persons at the initial phases of their training such as were represented in this study.

Results of testing Hypothesis 5. Hypothesis 5 states that the agreement between diagnosticians' judgments about patients and those made by the patients' psychotherapists would increase as a function of the amount of therapeutic contact the latter had with the patients.

For 10 of the 30 patients, criterion Q sorts were obtained from the therapists at a point in therapy well beyond the 10th hour, the point at which the original criterion Q sorts were routinely obtained. The number of therapy interviews which had taken place at the time of the terminal Q sort ranged from 15 to 49 (median of 23). In each case the clinician's final sort was correlated with the terminal criterion sort of the therapist and the distribution of such coefficients was obtained. The coefficients obtained between the clinical sorts and the 10th-hour sorts for the 10 cases were also noted and a comparison between the two sets of data is interpreted as reflecting the relationship between amount of therapeutic contact and therapist/diagnostician agreement.[9] The data in this analysis do not support the hypothesis as stated. The average of the validity coefficients of the clinicians' Q sorts was the same when compared with therapists' judgments based upon more extended therapeutic contact, as when compared with therapists' sorts based upon only 10 interviews.

Summary and Conclusions

The relative contribution of four commonly used kinds of clinical information to the description of the personality characteristics of male veteran psychiatric outpatients was studied.

Five clinicians studied the group of patients from a psychodiagnostic point of view, and performed four Q sorts of items describing the patients' personality characteristics and current psychiatric status. Increasing amounts of tests and other relevant data concerning the patient were made available to the clinicians sequentially. The data considered by the clinicians included (*a*) a four-page Biographical Data Sheet (BDS), (*b*) the patient's MMPI profile, (*c*) a Rorschach protocol (administered, scored, and interpreted by the clinician who performed the four Q sorts), and (*d*) a diagnostic interview conducted by the clinician. In each case, the sequence in which the clinician obtained and considered each of the above four kinds of data was specified. Although the BDS was always the first data-group considered and Q sorted, the other three kinds of data were arranged in the six possible sequences and each sequence was followed in one, and only one, diagnostic case per each of the five clinicians. A Q sort was performed after the clinician considered each kind of data, and the final sort represented the clinician's integration and interpretation of all the available data concerning the patient under consideration.

---

[9] The therapists' terminal criterion sorts were also compared with the corresponding 10th-hour criterion sorts. The resulting correlation coefficients ranged from .580 to .850 (median .734), which suggests that the therapists did not markedly alter their impressions of their patients (as reflected by the Q sort) between the 10th hour of therapy and the point at which the terminal sort was performed.

Each patient was assigned to a psychotherapist in the usual clinic manner, with the exception that the psychological test results and the usual written report were not made available to the therapist. After the patient had been seen by the therapist for 10 interviews (in three cases for only eight interviews) the therapist performed a criterion Q sort, describing the patient as he saw him on the basis of his therapeutic contact and whatever record material was available on him.

On the basis of the comparisons between the diagnosticians' and therapists' Q sort descriptions of patients, the following conclusions were drawn:

1. Although in some instances a stereotype personality description was the most accurate description of the patient obtained, bettering even the clinician's judgments made on the basis of any or all of the available test, history, and interview data, clinicians were generally able to describe psychiatric patients more accurately than did a single stereotyped description obtained by reference to the prevailing conceptions of the typical psychiatric patient seen in that setting.

2. Significant differences among the kinds of data were observed in terms of their contribution to accurate personality descriptions.

3. The diagnostic interview was consistently observed to contribute to greater accuracy of judgments about the personality characteristics of the patients. The interview was most useful when it was held early in the diagnostic sequence.

4. Neither the Rorschach nor the MMPI appeared to increase the validity of previous judgments in a significant number of cases (more frequently than they detracted from previously established validity of judgments).

5. Biographical data appeared to be a promising source of information in accurately assessing the personality characteristics of psychiatric patients. The BDS at least "held its own" relative to the other data even though underrepresented in the item pool and routinely considered without collateral information or data about the patients.

6. The relationship between the amount of information available to the diagnostician and the accuracy of his judgments is complex rather than linear, and varies according to the particular type (or kind) of data made available to the clinician.

7. Clinicians differ among themselves with respect to their over-all diagnostic accuracy, and also in terms of their success with the various tests or kinds of data utilized in the diagnostic process.

8. There was little difference in the degree of agreement between diagnosticians and therapists as a function of the amount of therapeutic contact (over a certain minimal amount) the latter had with specific patients.

9. The over-all agreement between diagnosticians and therapists was rather modest (mean $r = .48$), suggesting that somewhat different frames of reference were emphasized by each.

REFERENCES

American Psychiatric Association, Committee on Nomenclature and Statistics. *Diagnostic and statistical manual.* Washington, D. C.: Amer. Psychiat. Assoc. Ment. Hosp. Serv., 1952.

Block, J., & Bailey, D. *Q-sort item analyses of a number of MMPI scales scored on an Air Force officer sample.* Berkeley: Univer. Calif. Inst. Pers. Assess. Res., 1953. Pp. 1–27. (Dittoed)

Dailey, C. A. The effects of premature conclusions upon the acquisition of understanding of a person. *J. Psychol.,* 1952, **33**, 133–152.

Halbower, C. C. A comparison of actuarial versus clinical prediction to classes discriminated by the Minnesota Multiphasic Personality Inventory. Unpublished doctoral dissertation, Univer. of Minn., 1955.

Hathaway, S. R., & Meehl, P. E. The Minnesota Multiphasic Personality Inventory. *U. S. Depts. Army & Air Force, Milit. Clin. psychol. Rep.,* 1951, TM 8-242, AFM 160-45.

Kelly, E. L. Theory and techniques of assessment. *Annu. rev. psychol.,* 1954, **5**, 281–310.

Murray, H. A. *Explorations in personality.* New York: Oxford Univer. Press, 1938.

(Received December 24, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

A NOTE ON THE POSTURAL SWAY DISTRIBUTION[1]

FRED HEILIZER[2]

University of Rochester

The test most often used as a measure of hypnotizability is the postural sway test. Its reported correlations with various depth-of-hypnosis scales range from .50 to .75, with a probable mean of about .70 (Furneaux, 1952; Hull, 1933; Weitzenhoffer, 1953). The postural sway task involves direct repetitive suggestions to *S* that he is falling forward. (Much less frequently used are suggestions to fall backward or alternately sway forward and backward.) The suggestions are repeated for periods ranging from 30 sec. to 2½ min. The largest single forward excursion is usually taken as the score. The experimental period is sometimes preceded by a control period in which *S* is told to stand still. When both periods are utilized, the score is usually the difference between the maximum excursion in each period.

The resulting distributions are either truncated or U-shaped. These distributions are sometimes attributed to gross measurement procedures. Hull (1933), in reviewing the literature on several of the direct suggestibility tests, concludes:

> A judicious examination of all the evidence leads one to the conclusion, however, that the seeming deviations from the normal or bell-shaped distribution are more likely the result of defective scales for measurement than of any inherent constitutional tendency (p. 101).

Eysenck (1952), on the other hand, assumes that the postural sway test, as described above, has a threshold and a ceiling which transform an underlying normal distribution into an apparently U-shaped distribution.

Method

During the course of a larger study we utilized a more precise measurement procedure unrestricted by conditions of either a threshold or a ceiling. A control and an experimental period, of 2½ min. each, were utilized. The postural sway test was introduced to *S* as a test of motor activity, no mention being made of hypnosis. The *S* was directed to stand with feet together flat on the floor, arms folded, and eyes closed. She was told to stand there, relaxed, until further directions were given. She was also told not to be afraid of falling since, if she did fall, *E* would catch her. At the end of the control period, the following instructions were administered:

> Now, as you stand there, I want you to concentrate on what I'm saying, on my voice. Just stand there, relaxed, doing nothing in particular but listening to me. Blot everything else out of your mind, concentrate on my voice and, as you listen to my voice, you will start to sway forward, swaying forward, falling forward, falling forward. . . .

The words "falling forward" were repeated for 2 min. and 30 sec., with a reminder approximately every 30 sec. to "just listen to what I'm saying."

Permanent recordings of all forward and backward movements were made on a 3-to-1 step-down kymograph drum by means of a cord and pulley arrangement attached at one end to *S*'s collar and at the other end to a metal stylus. The 6-in. drum allowed a maximum sway of 12 in. forward and 6 in. backward to be recorded. Measurements of deviations from the baseline were made at 5-sec. intervals. The 30 measurements for each period were averaged, and the difference between

---

[1] This paper was abstracted from a dissertation submitted in partial fulfillment of the Ph.D. degree. I would like to express my gratitude to G. R. Wendt and E. L. Cowen for their interest and assistance in this study.

[2] Now at the VA Hospital, Northampton, Massachusetts.

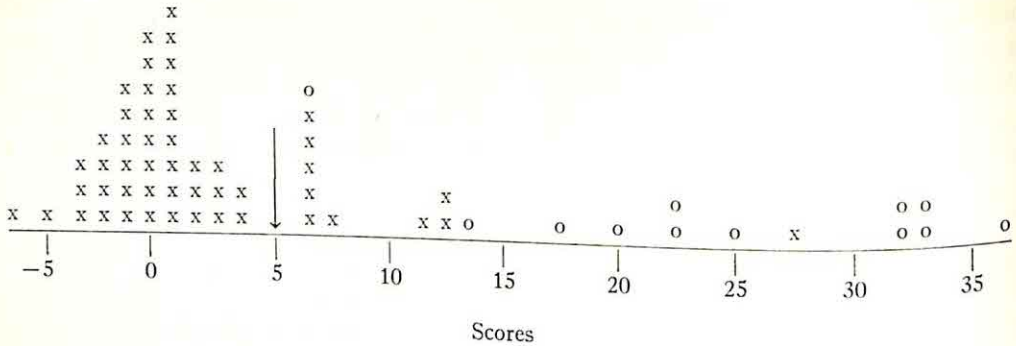


Fig. 1. Distribution of postural sway scores ($N = 62$). ("o" denotes *S*s who fell forward during the postural sway instructions.)

the two averages constituted the postural sway score. If *S* fell forward the maximum score of 40 1⁄10-in. units (corresponding to 12 in. of actual movement) was scored for the remainder of the 2½ min. The *S*s were 62 female college volunteers.

Results and Discussion

The resulting distribution is presented in Fig. 1. The numbers representing the class intervals are the midpoints of the intervals. The data, computed in tenths of an inch from recordings which were one third of actual size, are 3⅓ times as large as the actual movements or, conversely, the actual movements (in inches) are 30% as large as presented here. The distibution is markedly skewed to the right. The descriptive data for this distribution (total group) are presented in Table 1. The chi square indicates a significant departure from normality, while the *t* indicates that the sample differs significantly from zero.

Table 1

Descriptive Data for Postural Sway

	Total Group ($N = 62$)	Low Group ($N = 40$)	High Group ($N = 22$)
Mean	6.37	−0.15	18.21
SD	10.82	2.22	10.17
Median	1.34	0.03	16.14
Chi square ($df = 4$)	56.48*	3.15	—
t	4.60*	0.42	8.21*

* Indicates the .05 level or better.

A further examination of Fig. 1 reveals an unusual situation: the lower two-thirds of the distibution ($N = 40$), divided from the upper third of the distribution at the arrow (65th percentile), seems to be normally distributed about zero. The chi square for the low group (Table 1) is not large enough to reject the assumption of normality ($P = .60$), and the *t* is not large enough to reject the null hypothesis of no departure from a mean of zero ($P = .70$). The chi square test for departure from normality was not computed for the high group because of the small *N*. The *t* for this group indicates that the mean is significantly different from zero.

The above results raise the possibility that the postural sway distribution may be a composite of two distinct groups, one group demonstrating no response and the other group demonstrating a marked positive response. This interpretation derives from the facts that (*a*) the low group is normally distributed around zero, (*b*) there is a discreteness between the low group and the high group, and (*c*) the gap between the two groups cannot be considered to be an artifact of the scoring procedure in which the maximum score of 40 is assigned for the remainder of the 2½ min. after *S* falls (see Fig. 1).

The negative scores (greater forward sway in the control period than in the experimental period) offer another point of interest. Negative scores are often interpreted as a sign of negativism on the part of *S*, sometimes in an abnormal or morbid sense (Hull, 1933; Landis, 1932). There is no reason, however, to suspect such negativism in so many *S*s in this study, since they all freely volunteered for the experiment. Furthermore, the negative and positive scores appear to be on a continuum as part of the same, unresponsive distribution.

Conclusions

Although limited by the size of the sample and lack of replication with diverse samples, the above data indicate: (*a*) that the postural sway distribution is not normal and may well be composed of two separate groups, and (*b*) the interpretation of negativistic attitudes or personalities, made on the basis of negative postural sway scores, must be applied with care and only after additional criteria are met.

REFERENCES

EYSENCK, H. J. *Dimensions of personality*. London: Routledge & Kegan Paul, 1952.

FURNEAUX, W. D. Hypnotic susceptibility as a function of waking suggestibility. In L. M. LeCron (Ed.), *Experimental hypnosis*. New York: Macmillan, 1952. Pp. 115–136.

HULL, C. L. *Hypnosis and suggestibility*. New York: Appleton, 1933.

LANDIS, C. An attempt to measure emotional traits in juvenile delinquency. In K. S. Lashley, C. P. Stone, C. W. Darrow, C. Landis, & L. L. Heath (Eds.), *Studies in the dynamics of behavior*. Chicago: Univer. Chicago Press, 1932.

WEITZENHOFFER, A. M. *Hypnotism*. New York: Wiley, 1953.

(Received October 10, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

THE SET TO CLAIM UNDESIRABLE SYMPTOMS IN PERSONALITY INVENTORIES

CLINTON B. DESOTO AND JAMES L. KUETHE
Johns Hopkins University

The importance of subjects' (*Ss*') response sets in their performance on tests and questionnaires has been noted for many years. Cronbach (1950) discusses several response sets which depend on the form of test items, such as the tendency, varying in degree among *Ss*, to say "true" to items which may be answered "true" or "false," or "agree" to questions which may be answered "agree" or "disagree." This response set is known as the acquiescence set. Response sets seem to operate whenever items are difficult or ambiguous.

Most items on personality inventories appear ambiguous enough to arouse the operation of response sets. Items such as these from the Psychasthenia (*Pt*) scale of the MMPI, for example, fail to give *S* unambiguous standards for his report of his behavior and experiences:

> My memory seems to be all right.
> I certainly feel useless at times.
> I have strange and peculiar thoughts.

The phrases "all right," "at times," and "strange and peculiar" do not clearly specify the severity, frequency, and nature of symptoms which are required for endorsement of the items. The *Pt* scale consists of 48 items like these, which *S* must answer "(mostly) true," "(mostly) false," or "cannot say." Since 39 of the items are keyed positively, a high score on the scale could be due to a high degree of acquiescence on the part of *S* rather than to a high degree of what the *Pt* scale purports to measure (Hathaway & McKinley, 1951). Fricke (1956) recently pointed out that in most personality scales most items are keyed in one direction, making them potentially susceptible to serious distortion by this response set.

This article reports an exploration of the operation of response sets in the *Pt* scale of the MMPI.

Method

A balanced *Pt* scale was constructed in which 24 of the items are keyed positively and 24 negatively. This was accomplished by reversal of 15 of the 39 items which are normally keyed positively. These items were rewritten so that an answer of "true" has approximately the opposite meaning from an answer of "true" to the original item; in general, an answer of "true" becomes a claim of freedom from an undesirable symptom rather than a claim of the symptom. The 15 reversed items are presented in Table 1, together with

Table 1

Reversal of *Pt* Items

MMPI Booklet Number of Original Item	Reversal
32	I find it easy to keep my mind on a task or job.
67	I feel I am as happy as others seem to be.
76	Most of the time I feel happy.
86	I have enough self-confidence.
159	I understand what I read as well as ever.
182	I am quite sure of my mental health.
189	I feel pretty energetic most of the time.
217	I seldom find myself worrying about something.
266	I seldom become very excited.
304	In school I found it quite easy to talk before the class.
321	I am seldom embarrassed.
336	I am usually patient with people.
342	I usually remember what people say to me.
356	I have good powers of concentration.
361	I am inclined to take things in my stride.

their MMPI booklet numbers. These 15 items, together with the 9 negatively keyed items in the original *Pt* scale, constitute the 24 negatively keyed items in the balanced *Pt* scale. The remaining 24 positively keyed items in the original *Pt* scale were not altered in the balanced *Pt* scale.

The balanced *Pt* scale was given in printed form to 44 Johns Hopkins University summer school students in an introductory psychology course. These *S*s were of both sexes and varied widely in age and background.

The balanced *Pt* scale was also given in printed form to 95 Johns Hopkins winter students. Another 95 winter students took the original *Pt* scale in printed form. Like the summer school students, the last two groups were students in an introductory psychology course, but they were all males and were more homogeneous in age and background than the summer school group.

The "cannot say" response was made unavailable to the *S*s in these forms of the *Pt* scale, in order to permit fullest operation of the acquiescence set.

Results

One way of evaluating acquiescence in a scale is to correlate scores on the positively keyed items with scores on the negatively keyed items. Such correlations are reduced by acquiescence, sometimes even becoming negative, since acquiescence is positively related to scores on the positively keyed items but negatively related to scores on the negatively keyed items.

The positively keyed and negatively keyed subscales of the balanced *Pt* scale correlated .66 for the summer school students and .59 for the winter school students. Application of the Spearman-Brown formula in the manner customary for split-half correlations raises these figures to .80 and .74, which are scarcely lower than reliability coefficients typically reported for such scales (Edwards, 1957; Hathaway & McKinley, 1951).

The intercorrelations between the oppositely keyed halves of the balanced *Pt* scale are so high, in fact, that acquiescence set can be ruled out as an important factor in *S*s' performance on the scale. *S*s do not show tendencies to agree or to disagree with these items without regard to their content.

In tests for which acquiescence is an important factor, it has been found that *S*s show fairly high average acquiescence, despite large individual differences (Cronbach, 1950). Thus, another possible manifestation of acquiescence sets is a higher average score for items when they are positively keyed than when they are reversed and negatively keyed. Table 2 shows that the 95 *S*s who took the original *Pt* scale obtained a mean score of 3.57 on the 15 items which were reversed on the balanced *Pt* scale. The 95 *S*s who took the balanced scale obtained a mean score of 4.42 on these items as reversed. The difference between these means is significant at the .05 level by *t* test. Scores averaged lower on the positively keyed items than on the negatively keyed reversals. This result is the opposite from that which acquiescence would have caused. It is further evidence against the operation of acquiscence in *S*s' performance on these scales.

Other comparisons of the performance of *S*s on subscales of the original and balanced *Pt*

Table 2

Means and Standard Deviations of Scores on Subscales of the Original and Balanced *Pt* Scales

	Original Scale		Balanced Scale	
24 items positively keyed on both scales	*M*	5.07	*M*	5.59
	SD	3.61	*SD*	3.59
9 items negatively keyed on both scales	*M*	2.82	*M*	2.77
	SD	1.71	*SD*	1.56
15 items positively keyed on original scale, but reversed and negatively keyed on balanced scale	*M*	3.57	*M*	4.42
	SD	2.53	*SD*	2.70

scales, as shown in Table 2, reveal no significant differences. The distributions of scores as summarized by means and standard deviations are similar for corresponding subscales of the two scales. Even the one significant difference is actually rather small. It appears a reasonable conclusion that balancing the *Pt* scales did not appreciably influence *S*s' scores, either on the actually reversed items or, through context, on the unaltered items.

Discussion

It was remarked earlier that most items on the *Pt* scale are ambiguous enough that response sets would be expected to operate in *S*s' answers to them. Yet the results show that the acquiescence set, ordinarily a powerful factor when ambiguous true-false items are given, is essentially absent as a factor in the *Pt* scale. The *S*s failed to take the usual advantage of the items' ambiguities, so to speak. Why?

There is evidence that *S*s (nonpsychotic *S*s, at least) quickly notice that items like those on the *Pt* scale are concerned with personal ailments and troubles, with symptoms. Wiener (1948) found that the *Pt* scale is composed almost entirely of "obvious" items —"those to which significant responses were relatively easy to detect as indicating emotional disturbance" (p. 165). De Soto, Kuethe, and Bosley (1959) found that *S*s agree well in judging such items as to their desirability or undesirability for well-being, and that implications for well-being seem to be the most salient property of these items.

Thus the fact that these items are concerned with symptoms is obvious and important to the *S*s, so important apparently that they respond in terms of it despite the ambiguous standards. What the *Pt* scale measures, then, is their tendency to claim (or deny) undesirable symptoms, a tendency on which they vary greatly. This tendency should be regarded as a response set, even though it is tapped or activated only by items with content dealing with symptoms, since the ambiguities in the items and *S*s' tendencies to fake (Meehl & Hathaway, 1946) make its relationship to actual symptoms tenuous.

The symptom-claiming set seems to be prepotent among response sets. It seems to be the major factor in the *Pt* scale and other scales with many "obvious" items, reducing their validity as measures of other traits and producing correlations between scales intended to measure separate traits. For example, the *Pt* scale correlates .84 with the *Sc* scale (Hathaway & McKinley, 1951), it correlates .92 with the Taylor Scale of Manifest Anxiety (Eriksen & Davids, 1955), and it correlates −.84 with Edwards' Social Desirability Scale (SDS) (Edwards, 1957). These intercorrelations approximate the reliability coefficients typically obtained for such scales (Edwards, 1957; Hathaway & McKinley, 1951) indicating that the scales measure largely the same thing.

Those who have recognized a symptom-claiming set have tended to treat it as response sets are usually treated, that is, as a nuisance to be eliminated or controlled. The *K* scale was developed to correct scores on MMPI clinical scales for essentially this response set, and its authors remark that the *Pt* scale could be similarly used (Meehl & Hathaway, 1946). Edwards suggests that the SD Scale might serve better than the *K* scale to correct the MMPI clinical scales, and he further has developed a forced-choice personality test designed to prevent the operation of the response set (Edwards, 1957).

There are reasons for believing, however, that this symptom-claiming set is highly deserving of continued study as a personality variable in its own right.

Its prepotency over the otherwise powerful response set of acquiescence makes it appear important.

It seems to be the easiest of all personality variables to measure well with true-false items. This may be due to its prepotency as a response set, but in any case it argues further for its importance in people's psychological functioning. The correlations between the *Pt* scale and the *Sc*, Manifest Anxiety, and SDS mentioned earlier are extremely high in comparison with most correlations between measures of personality variables, and the intercorrelations of the last three scales are equally high (Brackbill & Little, 1954; Edwards, 1957). These four scales were constructed independently to measure different personality variables (although their items were drawn from the same large MMPI pool), yet they turned out to be essentially measures of the same variable, the symptom-claiming set. The resistance of the *Pt* scale to effects from the reversal of its items in the present study also attests to the relative ease of good measurement of this response set.

Despite the evident importance of the symptom-claiming set, however, its function in personality is far from well understood.

The SDS, *Pt, Sc,* and Manifest Anxiety scales, which appear to measure primarily this response set, have been correlated with many other behavioral measures, and some of these correlations may be useful in understanding the response set. Eriksen and Davids (1955) report clinical ratings by H. A. Murray of 20 *S*s on the trait of "optimism-pessimism," based on interviews of the *S*s and their autobiographies, which showed a rank-order correlation of −.90 with the *Pt* scale and −.87 with the Manifest Anxiety scale. Unfortunately, however, most correlations between these scales and other behavioral measures have been much lower than this. Some of these low correlations may be attributable to unique factors on which these scales seem to have differential light loadings (Hathaway & McKinley, 1951; Taylor, 1956).

It is probable, however, that the symptom-claiming set itself is determined by more than one psychological factor, each of which may account for some of the diverse correlates of the response set. There is convincing evidence that scores on these scales depend heavily on chronic anxiety or emotional responsiveness (Taylor, 1956), on tendency to use the defense mechanism of repression (Eriksen & Davids, 1955; Eriksen, Kuethe, & Sullivan, 1958), and on effort to make a good impression on others (Edwards, 1957; De Soto et al., 1959).

If these or another set of underlying personality factors jointly determine scores on the symptom-claiming set and hence its correlations with other variables, it appears likely that they can only be uncovered by studies of the joint relationships among several selected personality scales and other behavioral measures, as opposed to the pair-wise correlations between measures which have typified past research involving the response set.

This is not to say, however, that ordinary factor analysis is necessarily appropriate to the problem.

Suppose, for example, that the response set (x) is a function of two factors, actual degree of anxiety (y) and degree of anxiety which the person can express (z). And suppose that an individual's score on the response set equals the *lower* of his scores on the two factors, that is,

$$x = \text{the lower of } y \text{ and } z.$$

If an individual's actual degree of anxiety is lower than the degree of anxiety which he can express, as limited perhaps by repression and suppression, he will express his actual degree of anxiety, by claiming symptoms, and no more. On the other hand, if the degree of anxiety which he can express is lower than his actual anxiety, he will only reveal the expressible degree, no matter how much lower this is than his actual anxiety.

At least two of the implications of such a model receive empirical support. This model implies that, if the distribution of scores on the factors is symmetric, the distribution of scores on the symptom-claiming response set will be skewed upward, and this is the case for all of the scales which are good measures of the response set (Edwards, 1957; Hathaway & McKinley, 1951).

The model also implies that people scoring low on the symptom-claiming set are more heterogeneous with respect to both factors than people scoring high on the response set, who can so score only by being high on both factors. Low scorers on these scales have indeed been observed to be more variable than high scorers on other behavioral measures, which may be influenced by these factors, such as insight into determinants of their behavior (Eriksen et al., 1958) and paired-associates learning (Spence, Farber, & McFann, 1956).

In contrast to the above formulation, a factor-analytic formulation would be of the form

$$x = ay + bz$$

and would be clearly inappropriate.

If it is the factor on which a person has the lowest score that determines his symptom-claiming set, then the scales which are excellent measures of the response set are actually excellent measures of this factor. The problem will be to determine which factor is the one being measured for a given individual, and this can very likely be accomplished.

Summary

This article reports a study of response sets in the *Pt* scale. It was pointed out that most items in the *Pt* scale and similar personality scales are ambiguous enough to arouse the operation of response sets, in particular the acquiescence set, since the items are of a true-false form.

A balanced *Pt* scale was constructed by reversal of 15 positively keyed items to make a total of 24 positively keyed and 24 negatively keyed items. High positive correlations were obtained between the positively keyed and negatively keyed subscales. Reversal of the 15 positively keyed items to make them negatively keyed did not lower *S*s' mean score on them. These results indicate that acquiescence is not important in the *Pt* scale.

Instead, the *Pt* scale measures a prepotent response set: tendency to claim symptoms, which varies widely in degree among *S*s. This symptom-claiming set is tapped or activated by items with content dealing with symptoms. The *Pt* scale and other scales with which it correlates highly seem to be excellent measures of this response set.

The diverse correlates of these scales indicate that the symptom-claiming set is determined by more than one factor, but possibly in such a manner as to make factor analysis an inappropriate technique for their study.

REFERENCES

BRACKBILL, G., & LITTLE, K. B. MMPI correlates of the Taylor Scale of Manifest Anxiety. *J. consult. Psychol.*, 1954, **18**, 141–148.

CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, **10**, 3–31.

DE SOTO, C. B., KUETHE, J. L., & BOSLEY, J. J. A redefinition of social desirability. *J. abnorm. soc. Psychol.*, 1959, **58**, 273–275.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor Anxiety scale and the Hysteria-Psychasthenia scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135–137.

ERIKSEN, C. W., KUETHE, J. L., & SULLIVAN, D. F. Some personality correlates of learning without verbal awareness. *J. Pers.*, 1958, **26**, 216–228.

FRICKE, B. G. Response set as a suppressor variable in the OAIS and MMPI. *J. consult. Psychol.*, 1956, **20**, 161–169.

HATHAWAY, S. R., & MCKINLEY, J. C. *Manual for the Minnesota Multiphasic Personality Inventory.* (Rev. ed.) New York: Psychological Corp., 1951.

MEEHL, P. E., & HATHAWAY, S. R. The *K* factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *J. appl. Psychol.*, 1946, **30**, 525–564.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

TAYLOR, J. A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

WIENER, D. N. Subtle and obvious keys for the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1948, **12**, 164–170.

(Received October 14, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

EFFECT OF A TRAINING SCHOOL EXPERIENCE ON THE PERSONALITY OF DELINQUENT BOYS[1]

SYLVAN B. CADITZ[2]

Fort Worden Diagnostic and Treatment Center, Port Townsend, Washington

One of the most extensive programs of treatment and rehabilitation functioning in our society—training schools for delinquent children—has yet to be put sufficiently to the test of objective evaluation. These schools exist in great part to promote salutary changes in individuals whose behavior, and presumably underlying personality, are more or less socially unacceptable and require modification.

It is important to understand the essential personality differences between delinquent adolescents and the normal adolescent population in order to be better able to evaluate the character and direction of expected changes to be effected by the training school experience. Early representative studies of delinquents, utilizing the MMPI, are those by Capwell (1945), who demonstrated that delinquent girls were more maladjusted on all of the eight MMPI clinical scales than were nondelinquents, and by Monachesi (1953), who found differences on all scales except *Hs* and *Mf* between his various groups of delinquent boys and two control groups. A broad research program was reported by Hathaway and Monachesi (1953), in which 4048 ninth grade pupils in Minneapolis were given the MMPI. Two years later, the records were pulled on those who had come to the attention of the authorities as a result of delinquent behavior. Findings revealed significant MMPI differences between the predelinquent and 200 subsequently nondelinquent screened control records on many scales with *Pa, Sc, Ma,* and especially Scale *Pd,* positively associated with delinquent tendencies.

The only research data found which appear to bear directly on the effects of a training school experience as evaluated by objective psychological tests was Capwell's 1945 study of delinquent girls. She repeated, 4 to 15 months later, a battery of objective-type personality tests, including the MMPI, administered to 101 girls at a state training school and 85 nondelinquent high schools girls. The results suggest the delinquent group in many respects become more "normal" during their incarceration.

The present study is concerned with the differences in personality traits, as measured by objective tests, between delinquent and nondelinquent adolescent boys. The primary problem to be investigated is whether, as a result of a training school experience, the delinquents become more like the nondelinquent population.

Method

Subjects

Delinquent group. The delinquent group was drawn from 195 boys judged delinquent by the juvenile courts and transferred to the State Reception-Diagnostic Center at Chehalis, Washington. The 195 boys were all the boys admitted and tested at the Reception-Diagnostic Center over a period of five months, except for four cases of nonreaders and those boys admitted during a three-week period when testing was discontinued due to the depletion of test forms.

Ninety-one boys of this group were subsequently transferred to institutions other than the training school (a state operated school for younger, more immature boys, or a forestry camp) or were paroled directly to the community. The effect of this procedure is clearly to assign to the training school, from the incoming delinquent group, the older, more hardened delinquent offenders and individuals most in need of external sources of control.

---

[1] Condensed from a doctoral dissertation submitted to the University of Washington and done under the direction of Charles R. Strother.

[2] Now at the King County Juvenile Court, Seattle, Washington.

Of the 104 boys transferred to the training school, three were released earlier than the two-month minimum stay required for inclusion in this study (usually the result of circumstances external to the boy's adjustment, i.e., family leaving the state), five left under emergency conditions and could not be retested, and two arbitrarily refused to cooperate for the second testing. The 94 remaining boys who constitute the experimental group averaged 15.6 years old (range 14.1 to 17.7) and were retested after an average lapse of five months, 27 days (range 2 months, 16 days to 10 months, 23 days) of residence in the training school. Only four of these boys were at the training school less than four months.

Nondelinquent group. All boys from the sophomore classes of two high schools that were present the day of testing comprised the nondelinquent group. One of the schools, located on the fringe of a large metropolitan area (Tacoma, Washington), provided 48 cases who were retested 6 months, 13 days later. The other high school, which serves the local community of about 5,000 population, provided 55 cases who were retested after an interval of 4 months, 28 days. Three cases were lost from each of the two schools due to withdrawals and absences at the time of retesting. Thus the total high school population retested numbered 97.

Measures

Tests. The instruments chosen to evaluate personality factors in this study are the Minnesota Multiphasic Personality Inventory (MMPI) and the Edwards Personal Preference Scale (EPPS). Differential diagnostic studies involving the MMPI are prevalent in the literature and there are also a substantial number of cases where it has been used to evaluate therapeutic changes in various types of populations (Gallagher, 1953; Monachesi, 1953; & Scofield, 1953).

As an additional, independent measure of personality, the EPPS was selected. A recent article by Merrill and Heathers (1956) points up the complementary relationship between the MMPI, which evaluates personality structure, and the EPPS, which reflects the status of the motivational systems. Extensive norms for the EPPS on an adolescent population have been provided by Klett (1957).

Administration. The following two modifications of traditional MMPI administration procedure were made for all *S*s of this study:

First, all groups at the time of testing were instructed to answer "true" or "false" to all the items utilized, thus eliminating the "cannot say" (*?*) variable as a scoring category. This was done because prior experience had indicated that a sizable proportion of delinquent boys invalidate their tests by inordinately high "cannot say" scores. Rather than lose cases by overly high "cannot say" scores, it was felt better to evaluate poor test-taking attitudes through the other validation scales.

A second modification in the procedure was the use of the shortened form of the test as suggested by Cottle (1953). The first 366 items plus seven additional items included in the *K* scale were answered, allowing for scoring of all original scales with the exception of *Si*.

The administration of the EPPS followed that of the MMPI for all groups tested, with the instructions remaining essentially the same as printed on the test booklet.

Scoring. It was felt important for the purpose of this study to re-evaluate the practice of eliminating from the data those records with higher than usual *L, F,* and *K* scores. As this study was primarily oriented toward changes in personality test variables, it seemed vital that the scores of all individuals capable of complying with the test directions be included to insure a representative sample. The rejection of MMPI records with unusually high *L, F,* or *K* scores (as commonly practiced) would eliminate from the experimental group a distinctive and significant segment of the training school population. In their prediction study, Hathaway and Monachesi (1953) found that 37% of boys having invalid MMPI records subsequently became delinquent as compared to only 20% of boys having valid records. They state, ". . . *a priori* it would seem that one of the most important factors making for invalidity among the records from boys is also related to the commission of the delinquent acts" (p. 111). To exclude significant changes found in these " invalid" records upon retesting from this body of data could not easily be justified.

Background Variables

The survey of earlier studies in this area had revealed two background variables that appear related to the test scores of delinquents: extent of intactness of the family and antisocial history. Each member of the delinquent group was rated in each of the two areas by examination of official records, including the case history.

The family background in each case was considered unbroken if the natural parents were living together at the time of commitment and there were no previous extensive separations. Only 30 of the 94 delinquent cases, or 33%, were products of unbroken homes. The remaining background variable, the degree of antisocial behavior, was evaluated in terms of two levels:

Level I—Extensive delinquency: serious offenses committed over a period of time and indicated by involvement in more than three larcenies, burglaries, or personal assaults as reported in the case records.

Level II—Moderate delinquency: indicated by serious offenses of limited frequency or minor offenses. This group includes all boys not in Level I. These are boys with three or less serious offenses or involvement in such acts as runaway, window breaking, sexual exposure, fire-setting, etc.

Final tabulation revealed that 28, or about 30%, of the total cases were rated as Level I or having a history involving extensive, acting-out behavior, while the others, 66 in all, fell into the more moderate delinquency category.

Treatment of Data

The experimental hypothesis of greater similarity between the delinquent and nondelinquent boys following the training school experience was evaluated by comparing their test results at the first testing and again at the second testing. In order to better evaluate each group's contribution to the obtained difference values, a comparison was then made of the results of the first testing with that of the second for each of the groups.

The role of each of the background va-

Table 1

MMPI Means and Standard Deviations of the Delinquent Boys Contrasted with Nondelinquent Boys at the Time of the First and Second Testings

	First Testing					Second Testing				
	Del.		Nondel.			Del.		Nondel.		
Scale	*M*	*SD*	*M*	*SD*	*M*[a] diff.	*M*	*SD*	*M*	*SD*	*M*[a] diff.
L	4.14	2.40	3.80	2.67	.34	4.02	2.59	3.68	2.44	.34
F	10.14	6.89	8.72	6.92	1.42	8.15	5.55	7.10	6.26	1.05
K	14.48	4.67	13.60	4.53	.88	15.86	5.33	14.93	4.46	.93
Hs	14.13	4.23	13.75	4.81	.38	13.22	3.88	13.14	4.26	.08
D	20.57	4.44	19.36	5.18	1.21	18.97	4.35	17.84	4.78	1.13
Hy	20.60	4.40	19.44	5.05	1.16	20.00	4.24	19.86	4.54	.14
Pd	29.45	5.80	23.84	5.35	5.61**	28.74	4.07	22.56	4.94	6.18**
Mf	21.79	4.60	22.55	4.27	−.76	21.61	4.29	22.52	4.50	−.91
Pa	13.18	5.01	10.35	3.63	2.83**	11.61	3.32	9.97	3.44	1.64**
Pt	30.26	6.36	29.05	6.39	1.21	27.79	5.30	28.25	5.65	−.46
Sc	32.85	8.47	31.60	9.78	1.25	31.18	8.26	30.65	8.69	.53
Ma	23.30	4.97	21.65	4.73	1.65*	22.13	4.48	20.85	3.96	1.28*

[a] A difference in which the delinquent group is lower than the nondelinquent group is indicated by a minus.
* Significant at or beyond the .05 level of confidence.
** Significant at or beyond the .01 level of confidence.

riables, intactness of the family and extent of delinquent history, was evaluated by independently comparing each of the alternatives within each variable with the nondelinquent group at both the first and second testings.

Results

Difference Between Groups

Table 1 compares MMPI scores for the delinquent and nondelinquent groups at the time of the first testing and again at the second testing. Statistically significant differences on the *Pd* and *Pa* scales at the .01 level of confidence and *Ma* scale at the .05 level at the first testing are in accord with previous studies of comparable delinquent and nondelinquent populations. The profile of MMPI differences approximately duplicates Monachesi's (1953) data comparing a training school group with nondelinquents, except that the present study shows a statistically significant difference for the *Ma* scale whereas Monachesi's *Ma* difference fell somewhat short of significance.

Following an average training school stay of approximately six months for the delinquents and a comparable interim period for the high school group, the profile of differences remains similar to the one found at the first testing. It is interesting to note, however, that the difference value on the *Pd* scale is larger than found at the first testing, while the difference values on the *Pa* and *Ma* scales are lower but nevertheless remain statistically significant.

Table 2

Statistically Significant EPPS Means and Standard Deviations of the Delinquent Boys Contrasted with the Nondelinquent Boys at the Time of the First and Second Testings

Scale	Delinquents		Non-delinquents		
	M	*SD*	*M*	*SD*	M^a diff.
1st Testing					
Change	14.48	3.53	16.00	3.45	−1.52**
Consistency	8.66	2.87	9.45	2.48	− .79*
2nd Testing—None					

[a] A difference in which the delinquent group is lower than the nondelinquent group is indicated by a minus.
* Significant at or beyond the .05 level of confidence.
** Significant at or beyond the .01 level of confidence.

Table 2 contrasts EPPS scores of the two groups. At the first testing, of the 16 EPPS scales, only *change* and *consistency* showed

Table 3

MMPI Means and Standard Deviations of the First Testing Contrasted with the Second Testing for the Delinquent and Nondelinquent Boys

Scale	Delinquents					Nondelinquents				
	1st Testing		2nd Testing			1st Testing		2nd Testing		
	M	*SD*	*M*	*SD*	M^a diff.	*M*	*SD*	*M*	*SD*	M^a diff.
L	4.14	2.40	4.02	2.59	.12	3.80	2.67	3.68	2.44	.12
F	10.14	6.89	8.15	5.55	1.99**	8.72	6.92	7.10	6.26	1.62**
K	14.48	4.67	15.86	5.33	−1.38*	13.60	4.53	14.93	4.46	−1.33**
Hs	14.13	4.23	13.22	3.88	.91*	13.75	4.81	13.14	4.26	.61
D	20.57	4.44	18.97	4.35	1.60**	19.36	5.18	17.84	4.78	1.52**
Hy	20.60	4.40	20.00	4.24	.60	19.44	5.05	19.86	4.54	−.42
Pd	29.45	5.80	28.74	4.07	.71	23.84	5.35	22.56	4.94	1.28*
Mf	21.79	4.60	21.61	4.29	.18	22.55	4.27	22.52	4.50	.03
Pa	13.18	5.01	11.61	3.32	1.57**	10.35	3.63	9.97	3.44	.38
Pt	30.26	6.36	27.79	5.30	2.47**	29.05	6.39	28.25	5.65	.80
Sc	32.85	8.47	31.18	8.26	1.67	31.60	9.78	30.65	8.69	.95
Ma	23.30	4.97	22.13	4.48	1.17*	21.65	4.73	20.85	3.96	.80

[a] A difference in which the second testing mean is higher than the first testing mean is indicated by a minus.
* Significant at or beyond the .05 level of confidence.
** Significant at or beyond the .01 level of confidence.

significant differences between the training school and high school boys. At the second testing there are no significant differences on any of the 16 EPPS variables.

Test-Retest Differences within Groups

An evaluation of the relationship between the extent and direction of changes occurring in each group's test performance at the second testing as compared with the first and the subsequent effects upon the between group comparisons appears worth-while. In the case of variables which showed no significant change of differences, the mean scores may have remained essentially static or else there may have been parallel changes in the mean scores of both groups. On the other hand, when significant changes in difference scores were found, there may have occurred differential changes in either one of the two groups.

The MMPI. Table 3 presents a comparison of MMPI scores for the two testings. There are statistically significant reductions of mean scores for the delinquent boys on Scales *F*, *Hs*, *D*, *Pa*, *Pt*, and *Ma* and increase of *K*. The impression is unmistakably of movement in the direction of lessened personality disturbance and greater maturity.

A similar comparison of MMPI scores for the nondelinquent group shows a significant reduction of mean scores on the *F*, *D*, and *Pd* scales and an increase of the *K* score value. As in the case of the delinquent boys, the impression is also of movement in the direction of personality stability and integration.

EPPS. Table 4 presents a comparison of EPPS data from the first and second testings. The *affiliation* scores decreased significantly (at the .01 level of confidence) and the *aggression* scores increased significantly (at the .05 level of confidence) in the case of the delinquent group.

A similar comparison of EPPS scores for the nondelinquent group shows the only significant change occurring as an increase of the *autonomy* scores (significant at the .01 level of confidence).

The only two scales showing significant between-group differences at the time of the first testing (but not at the second testing), *change* and *consistency*, nevertheless do not demonstrate significant changes within each group.

Table 4

Statistically Significant EPPS Means and Standard Deviations of the First Testing Contrasted with the Second Testing for the Delinquent and Nondelinquent Boys

Scale	1st Testing		2nd Testing		
	M	*SD*	*M*	*SD*	*M*[a] diff.
Delinquents					
Affiliation	15.84	3.73	14.72	3.42	1.12**
Aggression	12.07	4.10	12.97	3.70	− .90*
Nondelinquents					
Autonomy	12.76	3.34	14.05	3.87	−1.29**

[a] A difference in which the second testing mean is higher than the first testing mean is indicated by a minus.

* Significant at or beyond the .05 level of confidence.

** Significant at or beyond the .01 level of confidence.

Family Intactness

Tables 5 and 6 present a comparison of MMPI data of the nondelinquent group with each of the two portions of the delinquent group divided on the basis of the intactness of the family.

First testing. Table 5 presents the MMPI scale comparison of the two delinquent groups with the nondelinquent group for the first testing. In addition to the *Pd*, *Pa*, and *Ma* scales, which are found to be significantly different for the total group (see Table 1), the broken home boys also produced significant difference on the *K* and *Hy* scales, whereas there is a decidedly reduced *Pd* difference and nonsignificant *Ma* difference when the unbroken home boys are compared with the high school boys. These results strongly suggest that an inordinate proportion of the personality pathology of the training school group is found in the test results of the boys from broken homes.

Second testing. A comparison of the MMPI results of delinquents from the two family backgrounds with the nondelinquents at the time of retesting is found in Table 6. Many of the indications of differential functioning of the two family background groups with respect to the high school boys were reduced (see Table 2), some items becoming statistically insignificant. The earlier reported in-

Table 5

MMPI Raw Means for Delinquent Boys of Two Family Backgrounds Contrasted with the Nondelinquent Boys at the Time of First Testing

Scale	Broken Home ($N = 64$)	Unbroken Home ($N = 30$)	Nondelinquent ($N = 97$)	D^a_{B-ND}	D_{UB-ND}
L	4.33	3.73	3.80	.53	−.07
F	10.58	9.20	8.72	1.86	.48
K	15.42	12.47	13.60	1.82*	−1.13
Hs	14.67	12.97	13.75	.92	−.78
D	20.83	20.03	19.36	1.47	.67
Hy	21.17	19.37	19.44	1.73*	−.07
Pd	30.70	26.77	23.84	6.83**	2.93*
Mf	21.72	21.93	22.55	.83	.62
Pa	13.38	12.77	10.35	3.03**	2.42**
Pt	30.91	28.87	29.05	1.86	−.18
Sc	33.61	31.23	31.60	2.01	−.37
Ma	23.88	22.07	21.65	2.23**	.42

[a] A difference in which the delinquent subgroup mean is lower than the nondelinquent group mean is indicated by a minus.
* Significant at or beyond the .05 level of confidence.
** Significant at or beyond the .01 level of confidence.

crease in mean differences of the two main groups on the *Pd* scale at retesting seems largely the result of increased scores on this scale by the unbroken family group of boys.

Seriousness of Delinquency Variable

Tables 7 and 8 present the test data of the training school group divided into two parts on the basis of seriousness of history of antisocial behavior, again with the expectation that there will be a difference in their movement toward normalcy. The serious offenders (Type I) and moderate offenders (Type II) are each compared independently with the high school group for both tests.

First testing. A comparison of the MMPI scores of each offense type with those of the nondelinquents is presented by Table 7. Essentially similar profiles are found which are in accord with the total group differences

Table 6

MMPI Raw Means for Delinquent Boys of Two Family Backgrounds Contrasted with the Nondelinquent Boys at the Time of Second Testing

Scale	Broken Home ($N = 64$)	Unbroken Home ($N = 30$)	Nondelinquent ($N = 97$)	D^a_{B-ND}	D_{UB-ND}
L	4.19	3.67	3.68	.51	−.01
F	8.81	6.73	7.10	1.71	−.37
K	16.13	15.30	14.93	1.20	.37
Hs	13.64	12.33	13.14	.50	−.81
D	19.20	18.47	17.84	1.36	.63
Hy	20.61	18.70	19.86	.75	−1.16
Pd	29.42	27.30	22.56	6.86**	4.74**
Mf	21.42	22.00	22.52	−1.10	−.52
Pa	11.75	11.30	9.97	1.78**	1.33
Pt	28.11	27.10	28.25	−.14	−1.15
Sc	31.47	30.57	30.65	.82	−.08
Ma	22.11	22.17	20.85	1.26	1.32

[a] A difference in which the delinquent subgroup mean is lower than the nondelinquent group mean is indicated by a minus.
** Significant at or beyond the .01 level of confidence.

Table 7

MMPI Raw Means for Delinquent Boys of Two Offense Types Contrasted with the Nondelinquent Boys at the Time of First Testing

Scale	Offense Type I ($N = 28$)	Offense Type II ($N = 66$)	Nondelinquent ($N = 97$)	$D^a_{\text{I-ND}}$	$D_{\text{II-ND}}$
L	4.14	4.14	3.80	.34	.34
F	9.29	10.50	8.72	.57	1.78
K	14.14	14.62	13.60	.54	1.02
Hs	14.57	13.94	13.75	.82	.19
D	21.43	20.21	19.36	2.07	.85
Hy	20.29	20.73	19.44	.85	1.29
Pd	30.32	29.08	23.84	6.48**	5.24**
Mf	21.46	21.92	22.55	−1.09	−.63
Pa	12.68	13.39	10.35	2.33**	3.04**
Pt	30.00	30.36	29.05	.95	1.31
Sc	32.29	33.09	31.60	.69	1.49
Ma	21.89	23.89	21.65	.24	2.24**

[a] A difference in which the delinquent subgroup mean is lower than the nondelinquent group mean is indicated by a minus.
** Significant at or beyond the .01 level of confidence.

(see Table 1), except that only the moderate offender is significantly different from the nondelinquent on the *Ma* scale.

Second testing. The MMPI differences for the two delinquency level groups and the nondelinquents at the second testing are given in Table 8. In general, the pattern of differences remains similar to those of the first testing. The *Pa* and *Ma* scores of the moderate offenders have decreased somewhat but remain significantly different from the high school boys, while the significant *Pd* mean difference has further increased. In the case of the more serious offenders, their mean *Pa* score has decreased although still significantly different than that of the nondelinquents, while their stay at the training school involved an increase in difference of *Pd* and *D* scores.

Discussion

The results of this study confirm the presence of important differences between the

Table 8

MMPI Raw Means for Delinquent Boys of Two Offense Types Contrasted with the Nondelinquent Boys at the Time of Second Testing

Scale	Offense Type I ($N = 28$)	Offense Type II ($N = 66$)	Nondelinquent ($N = 97$)	$D^a_{\text{I-ND}}$	$D_{\text{II-ND}}$
L	4.54	3.80	3.68	.86	.12
F	7.50	8.42	7.10	.40	1.32
K	16.64	15.53	14.93	1.71	.60
Hs	13.89	12.94	13.14	.75	−.20
D	20.25	18.42	17.84	2.41*	.58
Hy	20.25	19.89	19.86	.39	.03
Pd	29.25	28.53	22.56	6.69**	5.97**
Mf	22.29	21.32	22.52	−.23	−1.20
Pa	11.54	11.64	9.97	1.57*	1.67**
Pt	28.29	27.58	28.25	.04	−.67
Sc	30.89	31.30	30.65	.24	.65
Ma	21.54	22.38	20.85	.69	1.53*

[a] A difference in which the delinquent subgroup mean is lower than the nondelinquent group mean is indicated by a minus.
* Significant at or beyond the .05 level of confidence.
** Significant at or beyond the .01 level of confidence.

personality structures of delinquent and nondelinquent boys as reported by earlier studies. However, evidence for modification of these differences in the direction of normalcy by the delinquent group as the result of a training school experience is limited. The net impression derived from the analysis of MMPI data is that, although the delinquent group made notable improvements in many areas of personality functioning during their institutional experience and remained essentially unchanged in others, the nondelinquent group made similar changes.

Perhaps the outstanding feature of the EPPS comparisons between the two groups is the failure to find both pre and post differences on the great majority of the scales, many, if not all, which appear related to popular notions of characteristic aspects of delinquent activity. This instrument has already detected considerable differences between college and high school students (Klett, 1957) and males and females (Edwards, 1954). The conclusion that presents itself, therefore, is that the delinquent and nondelinquent boys of this study are, in fact, essentially similar in terms of most of their basic psychological needs as measured by the EPPS. Evidently, it is primarily in the mode of expression of their needs that they differ. The implication of this hypothesis is, for example, that to some extent, a similar aggressive motive may be expressed in the behavior of the boy competing on the football field, the boy stealing a car, the boy struggling for a higher school grade, and the boy striking a younger child. With respect to the quantitative aspects of the basic psychological needs, it appears that delinquent and nondelinquent adolescent boys as represented here are far more similar to each other than to girls or populations of other age ranges.

The occurrence of the expected pattern of MMPI differences between delinquents and nondelinqquents suggests that the withdrawal of the *cannot say* response to the test items had no significant effect upon the scale comparisons among the various groups. Nevertheless, the nature of the effects of this modification upon the individual scales is not known. This is especially true considering that the procedure of forcing answers to the test items for all subjects may have inflated the validity, and possibly other, scales beyond what would be ordinarily found for these populations. The possibility of systematic effects of the somewhat atypical test administration procedures as used here raising the height of the entire profile cannot be ruled out, and caution should be exercised when comparing these data to other findings, especially for diagnostic purposes.

The tendency of the *K* scale, in distinction to the other scales, to be higher at the second testing suggests the delinquent group became more similar to the normal college population and less similar to the adult population—possibly on the basis of developmental changes (McKinley, Hathaway, & Meehl, 1956).

It is perhaps interesting to note that the *Pd* scale of the MMPI was developed specifically as a measure of tendencies toward antisocial, acting-out behavior (Hathaway & McKinley, 1951), and more than any of the other scales has proved both sensitive and reliable in detecting differences between delinquents and nondelinquents. In this study, the unbroken family group of delinquents *increased* their differences with the nondelinquents between the first and second testing on this scale to the extent that the confidence with which the null hypothesis could be rejected was raised from the .05 to the .01 level. It appears that, in a sense, this group did not learn to be nondelinquent while at the training school; rather, there was a crystalization of the very antisocial proclivities that were no doubt instrumental in leading to training school commitment. If resistance to modification of antisocial tendencies found in delinquents from unbroken homes is related to the presence of relatively more firm identifications with the family and other social groups, the training school experience may act to weaken essentially positive identifications or else fail to weaken negative, antisocial identifications.

In general, of the delinquents, the boys from broken family backgrounds appeared to gain most from the stable, routine pattern of living which may have operated to compensate for deprivations that were no doubt important antecedents to their maladjustment. Thus, there is a tendency for the training school to stabilize and strengthen the personality organization of delinquents; at the same time, however, to have little effect upon, and possibly hinder in some cases, the assumption of socially responsible conduct that is expected of members of our society.

Summary

Ninety-four boys committed to a state training school took the MMPI and the EPPS prior to entry and again after an average institutional experience of approximately six months. A group of 97 unselected high school sophomore boys took the same tests and were retested approximately six months later.

The results of the first testing for the delinquent and nondelinquent groups were compared for significant differences on 12 MMPI scales and 16 EPPS scales, and the second test results for the two groups were similarly compared. The differences between the first and second test results for each group were also compared. Finally, each of the subgroups of the delinquents, based on broken or unbroken family background and severe or moderate delinquent history, was independently compared with the nondelinquents on the MMPI.

The findings indicate:

1. The training school experience was not effective in modifying the significant differences between the delinquents taken as a group and nondelinquents which were found before the delinquents entered the training school.

2. Delinquents and nondelinquents do not differ in their basic personality needs as measured by the EPPS.

3. Both groups showed improvement in terms of lower MMPI mean scores when their second test results were compared with the results of their first testing.

4. In the case of the delinquent subgroup comparisons, the boys from broken families were more different from the nondelinquents than were boys from intact families. However, the broken family delinquents benefited more from the training school experience—mainly in terms of better personality organization and stability. Evidence is limited that subgroupings of delinquents, based on severity of offenses, responded differentially to the institutionalization.

5. The variable considered as the most appropriate measure of specifically delinquent traits, the *Pd* scale of the MMPI, was not materially affected by the delinquents' training school experience, a finding most pronounced in the case of boys from unbroken family backgrounds.

REFERENCES

Capwell, Dora F. Personality patterns of adolescent girls. II. Delinquents and nondelinquents. *J. appl. Psychol.*, 1945, **26**, 289–297.

Cottle, W. C. The MMPI: A review. *Kansas Stud. Educ.*, 1953, **3**, No. 2.

Edwards, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corp., 1954.

Gallagher, J. J. MMPI changes concomitant with client-centered therapy. *J. consult. Psychol.*, 1953, **17**, 334–338.

Hathaway, S. R., & McKinley, J. C. Manual for the Minnesota Multiphasic Personality Inventory. New York: Psychological Corp., 1951.

Hathaway, S. R., & Monachesi, E. D. *Analyzing and predicting juvenile delinquency with the. MMPI.* Minneapolis: Univer. Minnesota Press, 1953.

Klett, C. J. Performance of high school students on the PPS. *J. consult. Psychol.*, 1957, **21**, 68–72.

McKinley, J. C., Hathaway, S. R., & Meehl, P. E. The *K* scale. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956.

Merrill, R. M., & Heathers, L. B. The relation of the MMPI to the Edwards PPS on a college counseling center sample. *J consult. Psychol.*, 1956, **20**, 310–314.

Monachesi, E. D. The personality patterns of juvenile delinquents as indicated by the MMPI. In S. Hathaway & E. D. Monachesi (Eds.), *Analyzing and predicting juvenile delinquency with the MMPI.* Minneapolis: Univer. Minnesota Press, 1953. Pp. 38–53.

Scofield, W. A further study of the effects of therapies on MMPI responses. *J abnorm. soc. Psychol.*, 1953, **48**, 67–77.

(Received October 23, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

THE EFFECT OF ITEM STYLE ON SOCIAL DESIRABILITY AND FREQUENCY OF ENDORSEMENT

ARNOLD H. BUSS
University of Pittsburgh

In recent years there has been an increasing emphasis on the variables that affect response to personality inventories. Endorsement of inventory items has been found to be affected by response set (Fricke, 1956), acquiescence set (Bass, 1955), defensiveness (Meehl & Hathaway, 1946), and social desirability (Edwards, 1953). The potency of such variables is illustrated by the correlation of .87 obtained between frequency of endorsement and social desirability (Edwards, 1953).

One approach to minimizing the effect of such variables is to build a "validity" scale into the inventory, such as the *K* scale of the MMPI. Another solution is to use a forced choice technique, as Edwards (1954) has done in matching items for social desirability. A third technique would be to manipulate items at the source: alter the manner in which the items are written in order to control the effects of response sets, social desirability, etc. If this approach is to be successful, it is necessary to discover the effect of item style on social desirability and frequency of endorsement. The purpose of the present study was to do just this in an attempt to determine whether item-writing might be used to control "nonpersonality" variance on inventories.

Method

Seven item styles were used. The first consisted of the unelaborated item with no "stylistic" component, e.g., "I sometimes get into fights." The second and third styles were: "I have trouble controlling (my tendency to) . . ." and "I can't help . . . ," both of which refer to a lack of inhibition or control. The fourth style was: "I am tempted to . . . ," which implies that there are impulses to engage in the activity but conflict over acting on them or not. "I am guilty about . . ." (or alternatively, "I am concerned about . . .") is self-explanatory. "Like most people . . ." appeals to a consensus. The seventh item style, "I must admit . . . ," implies an awareness of one's own behavior, and perhaps a tolerance of it, somewhat like an admission of minor vices.

There were 10 different item contents, all relating to aggression:

1. I sometimes get into fights.
2. When I get mad, I say nasty things.
3. I tend to get into arguments.
4. I tend to use strong language.
5. When I get angry, I sometimes pick up the nearest thing and break it.
6. I tend to be rude to people I don't like.
7. I sometimes have outbursts of temper.
8. When I don't like a rule, I tend to break it.
9. I tend to make fun of others.
10. I dislike many people I know.

These were the unelaborated items. They were also written in the six different styles just mentioned. The 10 contents and 7 styles yielded 70 items. The 70 items were presented to 70 college men and 92 college women, who rated them for social desirability. A 9-point scale was used, with social desirability maximum at 9 and social undesirability maximum at 1.

Next, the 70 items were broken up into seven 10-item inventories. No content was repeated within a single 10-item set. These inventories were administered in standard fashion to 714 college men and women. These *S*s were different from the ones who rated for social desirability. The number of times an item was answered "true" constituted its frequency of endorsement. Each *S* was administered only one 10-item inventory in order to avoid his responding to the same content twice. The *N* for each of the seven inventories was approximately 100, with a range of 95 to 106.

Results and Discussion

For the purposes of the present study it was not necessary to use psychological scaling techniques in order to derive social desirability values. Therefore, the means of the social desirability ratings were used, and these are presented in Table 1. The means in the bottom row of Table 1 are for style, regardless of content. The unelaborated style may serve as a base against which the effect of the various item styles may be compared.[1] The unelaborated style yielded the lowest social desirability values. The Can't Help and Like Most People styles were not significantly different from the unelaborated style, the *t* values being 0 and 1.98, respectively. The failure of the Can't Help style to increase social desirability is consistent with the notion that lack of control over one's behavior is socially undesirable. The failure of the Like Most People style is surprising, because on the face of it this style appeals to the same kind of consensus that is involved in social desirability.

The remaining four styles significantly raised the level of social desirability. For the Trouble Controlling, Must Admit, and Tempted styles the increases over the unelaborated mean for social desirability were small but significant, *t*'s being 3.78, 5.67, and 5.42. The Guilt style yielded an extremely large elevation of social desirability, and the *t* was 14.94.

These data indicate that it is possible to manipulate social desirability simply by writing items in different styles. Most of the styles yielded only small increases in social desirability, but the Guilt style was effective in changing ratings from social undesirability (a rating of less than 5) to social desirability (a rating of more than 5).

The frequencies of endorsement are presented in Table 2. Percentages were used because of the slightly unequal *N*s for the styles and contents. The means in the right-hand column of the table show that there are differences in endorsement of the contents, regardless of the item styles. The values range from 12.7% for "breaking things" to 67.7% for "temper." These differences are of considerable importance in evaluating the effects of nonpersonality variables on response to inventory items. They suggest that the content of the item is an important determiner

---

[1] Each style was compared with the Unelaborated style, using *t* tests for correlated means. For nine degrees of freedom, the .05 level of confidence requires a *t* of 2.26, and the .01 level of confidence requires a *t* of 3.25.

Table 1

Social Desirability Values for Different Item Styles

Content	Trouble Controlling	Must Admit	Can't Help	Guilty	Like Most	Tempted	Unelaborated	Mean
1. fights	3.0	3.3	3.0	5.0	3.4	3.3	3.2	3.5
2. say nasty	3.1	3.0	3.0	6.0	2.8	3.4	2.7	3.4
3. arguments	4.2	3.6	3.3	5.3	3.5	4.1	3.2	3.9
4. curse	2.7	3.4	2.5	5.5	2.8	3.4	2.7	3.3
5. break things	2.9	2.6	2.6	5.2	2.5	2.9	2.4	3.0
6. rude	3.5	3.0	2.4	5.8	2.5	3.1	2.5	3.3
7. temper	4.4	3.8	3.6	5.9	4.6	3.6	3.6	4.2
8. break rules	3.1	3.1	2.8	5.8	2.7	3.6	2.7	3.4
9. ridicule	2.9	2.9	2.5	5.8	3.1	2.8	2.6	3.2
10. dislike	3.3	3.1	2.7	5.1	2.6	3.0	2.8	3.2
Mean	3.3	3.2	2.8	5.5	3.1	3.3	2.8	3.4

Table 2

Frequency of Endorsement in Percentages for Different Item Styles

	Trouble Controlling	Must Admit	Can't Help	Concerned Guilty	Like Most People	Tempted	Unelaborated	Mean
1. fights	23	50	32	20	54	47	35	37.2
2. nasty	59	79	42	78	74	76	64	67.4
3. arguments	36	42	29	16	73	59	38	41.9
4. curse	18	51	15	29	60	56	38	38.1
5. break	10	12	5	24	5	22	11	12.7
6. rude	29	27	32	56	25	36	40	35.0
7. temper	55	77	54	46	90	75	77	67.7
8. rules	41	45	22	22	41	41	23	33.6
9. ridicule	28	77	15	40	33	34	27	36.2
10. dislike	21	22	10	14	23	32	12	19.1
Mean	32.0	48.2	25.6	34.5	47.8	47.8	36.5	

of endorsement, regardless of the effect of other variables. Contentual differences resulted in a range (of the means) of 55% in endorsement, whereas style resulted in a range of 21.6%. It is clear that if nonpersonality variables were to be controlled, there would still be considerable variance attributable to contentual (personality) variables.

The means in the bottom row of Table 2 reveal pronounced differences in the number of times an item is answered "True" as a function of item style. Again the unelaborated style is the base against which to compare the other styles. The Trouble Controlling, Guilt, and Can't Help styles all depressed frequency of endorsement, but the difference is significant only for the Can't Help style ($t = 3.89$). The fact that the Can't Help style depressed frequency of endorsement suggests that there is a reluctance to admit lack of control. The Trouble Controlling style also implies a lack of control but more in the sense of an individual struggling to maintain control rather than being helpless. Therefore, the depressing effect of the Trouble Controlling style is somewhat less than the effect of the Can't Help style.

The remaining three styles all significantly elevated frequency of endorsement, t values being: Must Admit 2.27, Like Most People 2.53, and Tempted 3.90. If the assumptions made earlier concerning these styles are true, then the tendency to endorse an item is enhanced by (*a*) an implied "generosity" in admitting minor crises, (*b*) an appeal to a consensus, and (*c*) emphasizing impulses rather than actions.

Thus, item style affects both social desirability and the frequency of endorsement. Not all the styles employed here produced significant changes, and some of the significant changes are perhaps too small to have practical value. However, only six styles (other than unelaborated) were used, and it seems probable that other styles might be more effective than some of the present ones. Certainly other styles should be tested. It is clear that style is a potent variable, as may be seen in the change from 26% to 48% endorsement with content held constant. Thus, the results suggest that it may well be possible to control "nonpersonality" variance at the source, i.e., when the item is written.

Summary

This study investigated frequency of endorsement and social desirability in relation to item style. There were 10 item contents, all concerning aggression. Each item was written in seven different styles: *Trouble Controlling, Can't Help, Feel Guilty About,*

Tempted to, Like Most People, Must Admit, and the unelaborated content. The 70 items (10 contents, 7 styles) were administered to college students, and, in addition, social desirability ratings were collected.

It was found that item style significantly affected both social desirability and frequency of endorsement. Social desirability was elevated significantly by the Trouble Controlling, Must Admit, Guilt, and Tempted styles. Frequency of endorsement was significantly depressed by the Can't Help style and significantly elevated by the Must Admit, Like Most People, and Tempted styles.

REFERENCES

BASS, B. M. Authoritarianism or acquiescence? *J. abnorm. soc. Psychol.,* 1955, **51**, 616–623.

EDWARDS, A. L. The relationship between the judged desirability of a trait and the probability that a trait will be endorsed. *J. appl. Psychol.,* 1953, **37**, 90–93.

EDWARDS, A. L. *Edwards Personal Preference Schedule. Manual.* New York: Psychological Corp., 1954.

FRICKE, B. G. Response set as a suppressor variable in the OAIS and MMPI. *J. consult. Psychol.,* 1956, **20**, 161–169.

MEEHL, P. E., & HATHAWAY, S. R. The *K* factor as a suppressor variable in the MMPI. *J. appl. Psychol.,* 1946, **30**, 525–564.

(Received October 25, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

SOCIAL DESIRABILITY AND MMPI PERFORMANCE: THE INDIVIDUAL CASE[1]

JAMES BENTLEY TAYLOR

Washington Public Opinion Laboratory, University of Washington

It has been shown that much test-taking behavior is influenced by group norms of social desirability. In a recent review, Edwards (1957) cites examples of this tendency, demonstrating the influence of social desirability values in personality inventories, Q sorts, and descriptive check lists. Two experiments have demonstrated this relationship to be true of MMPI items (Hanley, 1956; Rosen, 1956). These studies report correlations between *grouped* data: a mean social desirability rating, as given by a group, is correlated with the probability of an item being endorsed. A further question remains: will there be a high relationship between an *individual's* answers to MMPI items and the social desirability values he assigns to these items? The usual practice of grouping the data may serve to obscure a large amount of individual variance. It is quite possible that most individuals are little influenced in their answers by the social desirability values of MMPI items, and yet that the grouped data will show a strong relationship. In order to clarify this issue, it is necessary to investigate the correlations between social desirability values and item endorsements on an *individual* basis. This study reports such an investigation.

A further question is also investigated here: Do people who answer the MMPI items in a markedly pathological (i.e., socially undesirable) manner do so *in spite of* a normal understanding of social desirability norms? Or, do they answer deviantly as a function of deviant notions of social desirability? The data reported here allow a test of these two alternatives.

Procedure

Three types of social desirability were studied. The first of these types is referred to here simply as the "social desirability ratings"—the subject (*S*) was asked to tell how "desirable or undesirable" a trait was in others. The second is referred to as "imputed social desirability"—the *S* was asked to rate the social desirability of the items in the same manner as "most people" would. The third type of social desirability is referred to here as "normative social desirability"—the median social desirability rating for an item was computed, this value serving as "normative rating."

Seventy male adult schizophrenics, patients in two VA hospitals, served as *S*s. Ages of the *S*s ranged from 18 to 56, the median age being 33. The MMPI had been given to this group as part of a routine admissions battery. Within four months of the admissions testing, the patients were seen again and asked to rate 205 MMPI items on a nine-point scale of social desirability. These 205 items consisted of all statements in the *K*, *D*, *Pd*, *Sc*, and *SD* scales of the MMPI. The instructions were adapted from those used by Edwards (1957) in eliciting social desirability values. Two weeks later the same *S*s were asked to rate the same 205 items on the basis of imputed social desirability. Instructions were as follows:

---

[1] Adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy, University of Washington, 1958. The author wishes to thank Allen L. Edwards, Lillian Favero, William Fordyce, William Kogan, Clifford Lunneborg, and John Marks for their help in the completion of this research; he also wishes to thank the Computer Committee of the University of Washington for alloting computer time to this project.

> This is a test which tells how accurately you can "size up" the way other people behave.
>
> A large group of people, men and women, rated a number of statements. They rated each statement so as to show how desirable or undesirable it was in other people. An average rating of the desirability of each statement was made from these judgments.
>
> The mimeographed booklet contains all the statements that were rated. *Your task is to figure out how the average person rated each statement.* In other words, rate the statement like you think most people did. Put your rating to the left of the statement on the line provided. DO NOT SKIP ANY STATEMENTS; RATE EACH ONE. In making your ratings, use the scale shown below.

A 9-point rating scale was used.

Thus each patient rated the 205 MMPI items three times, first giving an item endorsement, secondly giving a social desirability rating, and thirdly rating the items for imputed social desirability.

Normative social desirability values were not based on data from the patient group. Instead, 80 evening class students at the University of Washington were asked to rate the MMPI items for social desirability. Median ratings for each item were computed; these served as the normative social desirability values.

Previous studies have usually correlated the median social desirability ratings as given by one group with the number of item endorsements given by another group. This design was followed here. The probabilities of item endorsement, as given by the schizophrenic group, were correlated with the median social desirability values (the normative ratings) given by evening class students. This correlation describes the relationship between grouped data.

Correlations on an *individual* basis were computed for each of the 70 patients. These correlations were between the following variables: (*a*) Item endorsement, (*b*) Social desirability scale values, (*c*) Imputed social desirability scale values, and (*d*) Normative social desirability scale values. Thus for each *S*, six correlations were obtained. These correlations were machine computed (Dvorak & Wright, 1956). In computing the correlations, item endorsement was a dichotomous variable, while the other variables were continuous. All correlations including item endorsement are point-biserial coefficients. In order to facilitate computation of mean and standard deviations, all correlations were transformed to Fisher's Z'.

Results and Discussion

For the grouped data, a correlation of .79 was found between the normative MMPI ratings and the probability of item endorsement. This value is similar in magnitude to those reported by Hanley (1956) and Rosen (1956) in their studies of MMPI endorsement and social desirability. But for the individual case, the relationship between item endorsement and social desirability was found to be relatively slight. This low correlation was true of all three types of social desirability studied: Imputed social desirability, Normative social desirability, and Social desirability. The mean correlations, computed on an individual basis, are presented in Table 1.

On the face of it, these findings suggest that grouping the data accentuates the relationship between social desirability and item endorsement. It would appear that we are not justified in assuming that social desirability accounts for much of the variance in any individual case. In fact, for the population studied, social desirability accounts for only about 10% of the variance.

This conclusion is qualified, however, by the differences in reliability of the data. The grouped data, based on means, undoubtedly has higher reliability than does a single individual's ratings or item endorsements. If we conservatively assume a reliability of .60 for a single individual's ratings and item endorsements, and make a correction for attenuation, the correlation between item endorsement and social desirability would be about .60. This approaches the value of the correlation between the grouped data. The present research gives no measure of individual reliability and so cannot evaluate the importance of the reliability problem to the findings.

A further qualification must be noted too. The item endorsements were often collected quite some time before the social desirability ratings were obtained. This time lag may have attenuated the relationship between individual item endorsements and individual ratings of social desirability.

Table 1

Mean Correlations of MMPI Ratings for 70 Schizophrenics

Correlation Between:	High Scorers Mean *r*	Low Scorers Mean *r*	Total Mean *r*	Number *r*'s Sig. ≤ .05 Level
Item endorsement Imputed social desirability	.21	.52**	.36	44
Item endorsement Social desirability	.19	.53**	.36	48
Social desirability Median social desirability	.56	.69*	.64	63
Imputed social desirability Median social desirability	.60	.67	.65	61
Social desirability Imputed social desirability	.54	.66	.62	62
Item endorsement Median social desirability	.11	.60**	.39	54

* Difference between high and low scorers significant at or beyond the .05 level.
** Difference between high and low scorers significant at or beyond the .01 level.

But in spite of these methodological problems, the findings clearly suggest that future research should differentiate between the influence of social desirability norms on grouped data, and the influence of these norms on the individual case. This research also argues against a simplistic interpretation of the MMPI as being, in large part, a measure of social desirability.

And what of the people who give markedly pathological answers to the MMPI items? Do they have deviant notions of social desirability? And if so, can their deviant test responses be explained as a function of deviant social desirability norms? In order to answer these questions, the 24 *S*s who gave the largest number of pathological responses were compared with the 24 who gave the lowest number of such responses. This data is also reported in Table 1.

As can be seen, the "high scorers" have slightly deviant notions of social desirability. This slight deviancy does not account for their endorsement of socially undesirable MMPI items. The "high scorers" are little influenced by their notions of social desirability when filling out the MMPI.

Summary

This study investigates the relationship between *individual* judgments of social desirability and *individual* item endorsements. Hospitalized schizophrenics endorsed MMPI items, rated them for "social desirability," and again for "imputed social desirability." "Normative social desirability" ratings were obtained from 80 University of Washington evening class students. Correlations between Item endorsement, Social desirability, Normative social desirability, and Imputed social desirability were computed for *each* of the 70 schizophrenic subjects. In addition, the mean probability of item endorsement for the schizophrenic group was correlated with the median social desirability rating given by the student group.

For the grouped data, the correlation was .79 between normative ratings of social desirability and probability of item endorsement, but only around .38 for the individual case. These findings militate against a simplistic interpretation of individual MMPI protocols as largely reflecting social desirability.

The data also permitted a test of the hypothesis that "people giving a large number of pathological (socially undesirable) MMPI responses give such responses as a function of deviant notions of social desirability." It was found that people who give many pathological responses do have slightly deviant notions of social desirability, but that this deviancy does not account for their endorsement of socially undesirable items. Rather, high scorers answer in a manner largely unrelated to their own notions of social desirability.

REFERENCES

DVORAK, A., & WRIGHT, C. *Symmetric correlation matrix program for the IBM Type 650.* (ONR Contract Nonr-477[08]). Seattle: Univer. of Washington, Div. Counsel. & Testing Serv., 1956.

EDWARDS, A. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

HANLEY, C. Social desirability and responses to items from three MMPI scales: *D, Sc,* and *K. J. appl. Psychol.,* 1956, **40**, 324–328.

ROSEN, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.,* 1956, **52**, 151–158.

(Received October 28, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

THE ATTITUDES OF MOTHERS OF JUVENILE DELINQUENTS TOWARD CHILD REARING

JEFF M. MADOFF

Connecticut State Hospital, Middletown

With the increasing awareness of the importance of the mother-child relationship as a determinant of child behavior, there has developed a strong interest in differentiating the child-rearing attitudes of parents of healthy children from those of pathological children. Though this kind of investigation has received some criticism (Orlansky, 1949; Symonds, 1937) as being too specific and circumscribed to truly measure the quality of the mother-child relation, its usefulness as a research technique has gained recent support. The present study was designed to determine if the attitudes of mothers of seriously delinquent adolescents (juvenile delinquents) were significantly different from the attitudes of mothers of healthy juveniles.

Though the study of parental attitudes has had a long and labored history, it has only been in recent years that reasonably objective techniques have been developed to assess the feelings of parents relating to child-rearing practices. The works of Shoben (1949) and Mark (1953) and Freeman and Grayson (1955) have substantiated the adequacy of easily administered attitudinal scales to differentiate the quality of the mother-child relationship among various disordered groups. In contrast to these studies, Greenbaum (1954), using the Shoben scale, found that maternal child-rearing attitudes were not related to the behavior of the child. L'Abate (1956) administered the Parental Attitude Research Instrument to parents of normal and problem children and obtained no differences in child-rearing attitudes for the two groups, though a differentiable personality configuration on the MMPI was established. Finally, Zuckerman, Oltean, and Monashkin (1958), examining mothers of schizophrenics and nonschizophrenics, were unable to relate severity of child-rearing practices to schizophrenia.

Method

Technique

The Parental Attitude Research Instrument (PARI) developed by Schaefer and Bell (1955) was the scale used to elicit information about child-rearing attitudes. This instrument, incorporating many items of the Shoben and Mark scales, is made up of 32 groups of items, each representing a fairly circumscribed area of parent-child behavior considered relevant to personality development. For purposes of research, the various scales may be combined in any number of ways according to theoretical concern, testing time limitations, specificity or generality of interest, and reliability. In order to tap a broad range of attitudinal areas and, at the same time, maintain an inventory of reasonable length and complexity, 20 scales were used in the present study. The scales of the PARI have been factor analyzed by Schaefer and Bell (1955) and Zuckerman, Ribback, Monashkin, and Norton (1958) with similar results. These factors have been extracted:

1. Factor A, Authoritarian-Control
2. Factor B, Hostility-Rejection
3. Factor C, Democratic Attitudes

The 20 scales used in this investigation are:

1. Comradeship
2. Suppression of Sex
3. Autonomy of Child
4. Encouraging Verbalization
5. Rejection of Homemaking Role
6. Abdication of Parental Role
7. Irritability
8. Marital Conflict

9. Strictness
10. Acceleration of Development
11. Approval of Activity
12. Deification
13. Martyrdom
14. Suppression of Aggression
15. Dependency
16. Intrusiveness
17. Excluding Outside Influences
18. Breaking the Will
19. Ascendance of the Mother
20. Seclusion of Mother

Subjects

The PARI was administered to 50 mothers of institutionalized delinquents and 57 mothers of healthy adolescents. The delinquents were selected from several residential agencies throughout the country on the basis of age, behavior history, intactness of home situation, and availability (for research purposes) of the mother. The normal youngsters were selected from three school systems known to be fairly representative of the general school population in terms of such variables as IQ distribution, academic achievement, and socioeconomic level.

The mean age of both the delinquent and the control sample was 14.4 years. The delinquents ranged from 12–17 years, while the normal youngsters had a range of 11–18 years. The delinquent group had a mean education of 7.7 grades, with a range of 4 to 10 grades; the normal *S*s had 8.5 years of schooling, with a range of 6 to 12 years. Within the delinquent sample there were 10 females and 40 males. Color composition was 42 whites and 8 Negroes. The normal sample contained 12 females and 45 males; 54 white youngsters and 3 Negroes.

The mean age of the mothers of the delinquents was 38 years, with a range of 31–47 years. No information as to age was available for the mothers of the control sample. The mothers of both groups were predominantly of grammar school education, with a mean of 10.6 years for the mothers of delinquents and 10.7 years for the mothers of the healthy sample. None of the mothers of the problem children attended college, while only one mother in the control sample had a college education.

In terms of religious background the two groups were somewhat different. The experimental group was essentially Protestant (30 of 34 members on whom such information was available); the control sample was made up of Catholics and Protestants in about equal number (23 Catholics and 27 Protestants).

On the basis of occupations reported by some of the mothers of each sample, it appeared that the normal sample represented a somewhat higher socioeconomic level than the delinquent sample. Within the delinquent sample, 13 working mothers were waitresses, four were domestics, and two were elevator operators. Among the working mothers of the normal sample, four were secretaries, four clerks, and two teachers. Most of the mothers of this sample did not work.

The two samples of mothers differ somewhat in terms of religious and socioeconomic background but are similar for the variables of education, color distribution, marital status, and family size. They are probably comparable in age too.

Results and Discussion

The two groups' responses were compared on each of the 20 scales of the PARI by means of the chi square model. On 7 of the 20 scales the mothers' responses proved to be significantly different at the .01 level of confidence. On two additional scales a significent difference was obtained at the .05 level of confidence. In *all* cases of significance the mothers of the delinquents expressed the more pathogenic child-rearing attitudes. The significant scales are presented in Table 1.

It is interesting to note that all nine of the differentiating scales had significant loadings on Factor A, relating to authoritarian-control behavior. In contrast to this, none of the scales of Factor B, Hostility-Rejection—or Factor C, Democratic Attitudes—differentiated between the two groups of mothers. It would appear, then, that the mothers of the disturbed children were substantially different from mothers of healthy youngsters in regard to attitudes relating to behavioral controls and the expression of authoritarian demands. Within this constellation of attitudes certain scales, when taken together, seem to afford a somewhat finer description of the differential quality of the mothers' perceptions of their roles in relation to their children. Thus

Table 1

Scales Significantly Differentiating Mothers of Delinquents From Mothers of Healthy Adolescents

Scale	Mothers of Normals		Mothers of Delinquents		χ^2	P
	Agree	Disagree	Agree	Disagree		
Suppression of Sex	10	47	28	22	15.59	.01
Deification	36	21	45	5	9.03	.01
Martyrdom	19	38	29	21	5.63	.01
Suppression of Aggression	13	44	26	24	8.58	.01
Fostering Dependency	16	41	34	16	15.72	.01
Excluding Outside Influence	22	35	37	13	12.14	.01
Seclusiveness	24	33	37	13	15.81	.01
Intrusiveness	21	36	29	21	4.01	.05
Breaking the Will	28	29	33	17	3.77	.05

the scales of Deification, Martyrdom, and Excluding Outside Influences as responded to by mothers of delinquents, seem to reflect the mother's perception of her role as one in which she sacrifices herself and her pleasures for her children and, in return, expects unquestioning loyalty, support, and devotion. The cluster of Suppression of Aggression, Dependency, and Intrusiveness seems to contain as its central theme the covert control of the child through keeping the child indebted to the mother, dependent and passive. This group of attitudes, as well as certain others, may relate to the concept of overpossessiveness as described by Levy (1943). Finally, the theme of harsh, punitive control somewhat related to the aforementioned parental dominance seems to filter through the scales of Breaking the Will and Suppression of Aggression. Together with the attitudes of overpossessiveness, these scales seem to reflect the total control of the child through overt dominance and covert control.

Summary

It was hypothesized that the child-rearing attitudes of mothers of juvenile delinquents are more pathogenic than are those of the mothers of normal adolescents. Fifty mothers of institutionalized juvenile delinquents and 57 mothers of healthy youngsters were compared on 20 child-rearing attitudes by means of the Parental Attitude Research Instrument. The two groups of mothers differed significantly on 9 of the 20 scales. In all cases of significance, the mothers of the delinquents expressed the more punitive, controlling, and authoritarian attitudes.

REFERENCES

Freeman, R., & Grayson, H. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **50**, 45–52.

Greenbaum, W. Influence of certain maternal attitudes on the behavior of rejected children. *Dissertation Abstr.*, 1954, **14**, 2124.

L'Abate, L. The personality of problem parents. Paper read at APA, Chicago, 1956.

Levy, D. *Maternal overprotection.* New York: Columbia Univer. Press, 1943.

Mark, J. The attitudes of the mothers of male schizophrenics toward child behavior. *J. abnorm. soc. Psychol.*, 1953, **48**, 185–189.

Orlansky, H. Infant care and personality. *Psychol. Bull.*, 1949, **46**, 1–48.

Schaefer, E., & Bell, R. Parental attitude research instrument. Unpublished manuscript, Laboratory of Psychology, National Institute of Health, 1955.

Shoben, E. The assessment of parental attitudes in relation to child adjustment. *Genet. psychol. Monogr.*, 1949, **39**, 101–148.

Symonds, P. Some basic concepts in parent-child relationships. *Amer. J. Psychol.*, 1937, **50**, 195–206.

Zuckerman, M., Oltean, Mary, & Monashkin, I. The parental attitudes of mothers of schizophrenics. *J. consult. Psychol.*, 1958, **22**, 307–310.

Zuckerman, M., Ribback, Beatrice, Monashkin, I., & Norton, J. Normative data and factor analysis on the parental attitude research instrument. *J. consult. Psychol.*, 1958, **22**, 165–171.

(Received October 30, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

SEXUAL SYMBOLISM, PERSONALITY INTEGRATION, AND INTELLECTUAL FUNCTIONING[1]

RUDOLPH MOOS AND PAUL MUSSEN
University of California, Berkeley

There have been a number of recent studies evaluating the psychoanalytic theory of sexual symbolism (Jacobs, 1954; Jones, 1956; Levy, 1954; Starer, 1955). Their findings support the hypothesis that, for adults at least, elongated, pointed or penetrating objects—and the words signifying such objects—are likely to be designated as male, while hollow, rounded, or enclosing objects are likely to be designated as female. Both psychiatric and normal subjects respond consistently with the sex symbolism hypothesis to a degree significantly greater than chance expectancy (Jacobs, 1954; Jones, 1956; Starer, 1955). The responses of normals, however, reveal greater agreement with the hypothesis than the responses of psychiatric patients (Jacobs, 1954; Jones, 1956). On the basis of this latter finding, it has been maintained that "severe impairment of personality integration is accompanied by a disorganization of symbolic processes" (Jones, 1956, p. 189). Observation of the performances of psychotics during a symbol test led one investigator to suggest the "possibility that the inability of certain individuals to adapt to and accept cultural symbolism results in emotional disturbance" (Starer, 1955, p. 454).

Such hypotheses can be substantiated only if variables such as age, education, and intellectual functioning are controlled. While the normal and psychiatric subjects in one of the studies (Jones, 1956) were matched for age and education, they were not matched for intellectual functioning. It is therefore possible that, in spite of the educational equality of the groups, the psychiatric subjects were not of the same level of intellectual functioning as the normals. If this is so, their failure to respond as adequately as the normal controls may be regarded as a reflection of general intellectual inefficiency rather than of a failure to accept the generally prevailing symbolism of the culture or a "disorganization of symbolic processes" (Jones, 1956) per se.

The purpose of the present study was to determine whether or not the responses of normals to sex symbols are more frequently consistent with the psychoanalytic hypothesis of sex symbolism than are the responses of psychiatric patients of the same age, educational background, and level of intellectual functioning.

Method

Subjects

There were three groups of 10 male *S*s each. The psychotic group had been diagnosed as schizophrenic and had been hospitalized in a Veterans Administration hospital for at least one year. The neurotic group consisted of patients who had been referred to the psychological service of a Veterans Administration clinic and had been diagnosed as neurotic. The normal group was made up of nonpsychiatric surgery patients at a Veterans Administration hospital.

*S*s were selected in such a way that the three groups were equated as closely as possible on the variables of age, education, and intellectual functioning. The last variable was evaluated on the basis of raw score on the 40-word vocabulary test of the Shipley-Institute of Living Test (Shipley, 1946). The results of the matching of the three groups are given in Table 1.

---

[1] This research was conducted while the senior author was a psychology trainee at the Veterans Administration Hospital, San Francisco, California. The authors wish to express their appreciation to Paul McReynolds and Lex Milton of the Veterans Administration Hospital, Palo Alto, California, for their cooperation in this study.

Materials

*S*s were presented with a test booklet consisting of three test sheets and an instruction page. The first test sheet consisted of a list of 40 words (20 "male" and 20 "female," listed in random order) taken from Jacobs's list of 113 words which, according to Freud's writings (1938), symbolized sex organs.

The second sheet contained, in random order, 10 elongated or pointed (male) figures and 10 hollow or rounded (female) figures. Some of these were taken from Levy's study (1954), and some from McElroy's (1954); others were specifically designed for this test. The third test sheet contained the 40 words of the Shipley-Institute of Living Vocabulary Test.

The instructions, presented in writing and orally, were similar to those used by Jones (1956). *S*s were asked to indicate, by writing the letters M or F by each word or design, "whether you think it is more like a man or more like a woman."

Results

For all three groups of *S*s, and for both the "figures" and "words" tests (considered separately), the proportion of responses consistent with the sex symbolism hypothesis was significantly greater than the chance expectancy of 50%. Thus, these data were consistent with the earlier findings that supported that hypothesis (Jacobs, 1954; Jones, 1956; Starer, 1955).

Table 2 presents the means and standard deviations for all three groups on the two tests. Scores were the number of responses consistent with the psychoanalytic sex symbolism hypothesis.

Table 1

Mean Age, Mean Number of Years of Schooling, and Mean Shipley Vocabulary Scores of the Three Groups of Subjects

Group	Mean Age	Mean Years of Schooling	Mean Vocabulary Score
Normals	36.5	10.9	27.5
Neurotics	36.7	11.5	27.5
Psychotics	36.3	11.6	27.6

Table 2

Means and Standard Deviations of All Groups on Words and Figures Tests

	Words		Figures	
	Mean	*SD*	Mean	*SD*
Normal	25.0	12.0	14.9	3.3
Neurotic	24.2	28.4	14.7	10.0
Psychotic	22.5	8.3	13.8	3.8

A simple analysis of variance of the data from both tests failed to reveal any significent differences among the three groups. Thus, our results are not consistent with earlier findings that normal *S*s respond to symbols (words or figures) in ways consistent with the sex symbol hypothesis to a significantly greater extent than psychiatric *S*s.

Discussion

The discrepancy between these and earlier findings may be largely explained by the fact that, in this study, unlike the earlier ones, normals and disturbed *S*s were matched for intellectual functioning. Thus, in the earlier studies, the effects of personality disintegration may have been confounded with the effects of intellectual inefficiency or deterioration which are frequent concomitants of severe emotional disturbance. It was therefore possible that the relatively poor performance of psychiatric patients on symbolism tests (i.e., their relative lack of agreement with the sex symbolism hypothesis) was a function of poorer intellectual functioning rather than of their psychiatric status or their level of personality integration. The data of the present study lend support to this argument, since if emotional disturbance or personality disintegration were the basic factor underlying the psychiatric patients' performance, the scores of these *S*s would have differed significantly from those of the normals.

Apparently, groups composed of individuals matched for intellectual functioning are likely to be equally accepting of symbols prevalent in the culture, regardless of their level of "personality integration." This finding suggests that the processes involved in interpreting sex symbols are related to *other cognitive* functions—intellectual functioning in particular—not to emotional status alone. In this connection, it is interesting to note that Starer had informally observed that "patients who were relatively more confused and disorganized [i.e., were functioning more poorly] made many more errors in matching than patients who seemed in fairly good states of remission" (1955, p. 454).

In brief, in view of the present data, statements linking restricted acceptance of the prevailing cultural sex symbols to personality disintegration seem overgeneralized. There was no support for the hypothesis that "one source of illness is the fact that there has never been a complete acceptance on the part of the very disturbed patient of the generally prevailing symbolism in a given culture" (Starer, 1955, p. 454). Moreover, these findings suggest that Jones's conclusion that "severe impairment of personality integration is accompanied by a disorganization of the symbolic processes" (1956, p. 189) is too extreme. It may be more reasonable to hypothesize that severely emotionally disturbed patients *of low intellectual functioning* are likely to show "disorganization of symbolic processes."

Summary

The purpose of the study was to determine whether or not normals, neurotics, and psychotics, matched in age, education, and level of intellectual functioning, differ in the extent to which their responses to sex symbols are consistent with the psychoanalytic hypothesis of sex symbolism. Thirty male subjects—10 normals, 10 neurotics, and 10 psychotics—designated each of 40 words and 20 figures as male or female. The vocabulary test of the Shipley-Institute of Living Scale provided the measure of intellectual functioning.

For all three groups, for both words and figures, the proportion of responses consistent with the sex symbolism hypothesis was significantly greater than chance expectancy. However, the three groups did not differ significantly from each other in the number of responses consistent with the hypothesis. These results support the theory of sexual symbolism, but they give no support to the generalization that emotional disturbance is accompanied by disorganization in the symbolic processes. According to the data of the present study, normal and psychiatric subjects who are equal in intellectual functioning do not differ significantly in the degree of acceptance of the prevalent cultural sex symbolism.

REFERENCES

FREUD, S. *The basic writings of Sigmund Freud.* New York: Modern Library, 1938.

JACOBS, A. A study of the responses of normals and mental hospital patients to Freudian sexual symbols. *J. consult. Psychol.*, 1954, **18**, 454.

JONES, A. Sexual symbolism and the variables of sex and personality integration. *J. abnorm. soc. Psychol.*, 1956, **53**, 187–190.

LEVY, L. H. Sexual symbolism: A validity study. *J. consult. Psychol.*, 1954, **18**, 43–46.

McELROY, W. A. A sex difference in preference for shapes. *Brit. J. Psychol.*, 1954, **45**, 209–216.

SHIPLEY, W. C. *Shipley-Institute of Living Scale for Measuring Intellectual Impairment: Manual.* Hartford, Conn.: Institute of Living, 1946.

STARER, E. Cultural symbolism: A validity study. *J. consult. Psychol.*, 1955, **19**, 453–454.

(Received October 31, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

WEIGHT PROBLEMS AND PSYCHOLOGICAL FACTORS[1]

JOSEPH FRIEDMAN
Temple University

In recent years, numerous articles have appeared which deal with the psychological correlates of weight problems and, more specifically, obesity. Various investigators (Bett, 1951; Brosin, 1953; Bruch, 1941; Conrad, 1954; Rennie, 1940; Richardson, 1946), working largely with incidental samples seen in clinical settings, have suggested that there are personality factors which are related to the weight problem. In short, weight deviancies are viewed as problems in psychosomatic relationships.

A host of traits have been ascribed to overweight people, but with few validity studies. Among these, increased anxiety and orality, lack of sociability, maternal overprotection and rejection, as well as a tendency for this problem to occur more frequently among only and youngest children are most often cited. At the other extreme, the problem of the underweight person has been relatively unexplored. Few studies bearing directly upon this phase of the problem are known to the writer.

The purpose of the present study was to determine possible psychological differences between college women identified as overweight, underweight, and normal weight. A list of variables which reflect the literature on this problem was constructed. Measures tapping these and other factors were selected. The three groups were compared with special attention to the following variables: (*a*) anxiety level, (*b*) emotional stability, (*c*) energy level, (*d*) impulsivity, (*e*) social interest, (*f*) social ascendency, (*g*) agreeableness, (*h*) cooperativeness, (*i*) reflectiveness, (*j*) subjectivity, (*k*) acceptance of responsibility, (*l*) femininity, (*m*) orality, (*n*) degree of positive parental relationships, and (*o*) birth order.

Method

Criteria and Definitions

Basic to this study is the determination of what constitutes overweight and underweight. The following criteria were set up for selection of the weight groups:

(*a*) Overweight Group (OWG): those persons whose actual body weight was 10% or more above the statistical mean for their age, sex, and height.

(*b*) Underweight Group (UWG): those persons whose actual body weight was 10% or more below the statistical mean for their age, sex, and height.

(*c*) Normal Weight Group (NWG): those persons whose actual body weight deviated from the statistical average for their age, sex, and height by no more than 6% in either direction.

The tables used to delimit the above groups were those reported by Dublin and Marks (1938). All weights in the present study were taken while wearing indoor clothing and shoes, as were the weights from which the reference tables were constructed. The weighing instrument was a bathroom scale which had been checked for accuracy.

To further reduce errors of assignment of *S*s to groups, it seemed best to omit borderline cases. Thus, weight deviations greater than 6% but less than 10% in either direction were not included in the sampling.

Selection of the Sample

A questionnaire, including items on age, height, and weight, as well as other information (in the nature of distractors), was administered to 476 college women. One hundred of these women were tentatively assigned to the three weight groups in accordance with their reported data. After interviewing these persons to gain their cooperation, they were actually weighed. Those who continued to meet our criteria were included in the sample. In this manner three groups, each containing 26 *S*s, were established.

---

[1] This study was derived from a dissertation submitted to the Temple University Graduate Council in 1958, in partial fulfillment of the requirements for the PhD degree. The writer wishes to express his appreciation to F. H. Lund, J. D. Page, and H. C. Reppert for their guidance and direction.

Description of the Sample

The mean ages (in years) of the three groups were: OWG = 18.69, UWG = 20.12, and NWG = 20.15. The differences between the groups were not significant.

The OWG had an average positive weight deviation of 24.8%, and the UWG had an average minus deviation of 13.1%.

Testing Phase

The Cattell Anxiety Scale, the Guilford-Zimmerman Temperament Survey, and the Gordan Personal Profile were administered to small groups of *S*s. The Blacky Test (Cards I and II) and the TAT (Cards 6GF and 7GF) were administered individually at a second session. All measures were evaluated in accordance with their respective manuals of instruction.

Results

Analysis of variance (single classification) of the self-report measures revealed only two variables which differed significantly between the groups. These were social ascendency as measured by both the Guilford-Zimmerman Temperament Survey and the Gordan Personal Profile ($F = 5.54$, $P < .01$ and $F = 6.95$, $P < .01$, respectively) and subjectivity also measured by the Guilford-Zimmerman scale ($F = 3.50$, $P < .05 > .01$). To isolate the source of these differences, the fiducial limits of the differences between the means were calculated (Table 1).

Table 1

Fiducial Limits Tests Applied to the Significant Factors

Factor	$\bar{x}-\bar{x}$	Difference	P
Ascendency (Guilford-Zimmerman scale)	NWG–OWG (17.69–17.81)	0.88	>.05
	NWG–UWG (17.69–13.70)	3.99*	<.01
	OWG–UWG (17.81–13.70)	4.11*	<.01
Ascendency (Gordan Profile)	NWG–OWG (56.12–55.35)	0.77	>.05
	NWG–UWG (56.12–51.23)	4.89**	<.01
	OWG–UWG (55.35–51.23)	4.12**	<.05, >.01
Subjectivity (Guilford-Zimmerman scale)	NWG–OWG (19.62–16.35)	3.27***	<.05, >.01
	NWG–UWG (19.62–16.42)	3.20***	<.05, >.01
	OWG–UWG (16.35–16.42)	0.07	>.05

* $t_{.05} = 2.78$, $t_{.01} = 3.69$.
** $t_{.05} = 3.18$, $t_{.01} = 4.22$.
*** $t_{.05} = 2.83$, $t_{.01} = 3.76$.

The results indicate that the UWG was less ascendent than the other two groups on both measures. With regard to subjectivity, both the OWG and the UWG were more subjective and hypersensitive than the NWG.

The projective data were independently rated by two clinical psychologists as to the degree of oral eroticism, oral sadism, and positive maternal and paternal relationships. The obtained reliabilities (ϕ coefficient) were as follows: positive maternal relations = .78, positive paternal relations = .67, and oral sadism = .47. All of these were significant at the .01 level of confidence. Oral eroticism was less reliable and did not attain significance. This factor was considered too unreliable to warrant further statistical treatment. Analysis of variance of the remaining three variables revealed no significant differences among the three weight groups.

In order to assess the relation of birth order to weight problems, the chi square technique was used. Only and youngest children from our sample were combined and compared against all others in the three weight groups. The resultant chi square value was 4.5099 ($df = 2$) and did not approach significance. Thus, birth order was assumed to be distributed throughout the weight range in accordance with chance variation.

Discussion

Prior to a discussion of the findings, two limitations of the present study must be stressed. The first of these is the small number of cases studied. Secondly, the magnitudes of the interrater reliabilities, while significant, are lower than is desirable. Thus, the negative results could conceivably be due to errors of measurement and to the small sample size. These points must be considered throughout the following discussion.

Two factors yielded significant differences between the three groups. These were: (*a*) less social ascendence in the underweight group as compared to the other two groups and (*b*) less objectivity and more hypersensitivity among both the overweight and underweight groups than in the normal weight groups.

These findings may be variously interpreted. From the classical psychoanalytic viewpoint, the submissiveness of the thin persons might be attributed to deprivation of social contact in the oral stage of infancy. However, such a deprivation might be expected to produce greater oral sadism and/or poor parental relationships as well. The present data do not support such contentions. Psychologists inclined toward the body typology theories might consider this finding to be in support of the ideas of Kretschmer (1925) and Sheldon and Stevens (1942). However, according to these theories, the overweight persons should also be more socially outgoing than the middle group. So far as the present study is concerned there is nothing to support this view. On the other hand, it would appear more tenable to suggest that lack of ascendency on the part of the underweight group reflects the subordinate social role which the individual of limited physique has been forced in many instances to assume. This, rather than factors of orality or inherent temperament, would be sufficient to account for the variable.

The second significant result of this study indicated that both extreme weight groups were less objective in thinking and more hypersensitive than the normal weight group. Here, too, the author suggests an interpretation consistent with learning theory. Either extreme of weight presents a social problem and opens one to social disfavor. Such pressures would tend to foster hypersensitivity as regards physique and general appearance. It may also be mentioned that present results are similar to those reported by Seltzer (1946) who found that any bodily disproportion was related to less stable and less well integrated personalities among college males.

In general, the results of this investigation tend to refute some widely held concepts. Beyond the two exceptions already mentioned, there were no psychological factors which distinguished the three weight groups to a significant degree. Studies which have provided a measure of support for the above include those of Laplane and Etienne (1953), Iversen, Juel-Nielson, Quaade, Tolstrup, and Ostergaard (1951), Iverson (1953), and McHale (1926) with children, and Young, Beresford, and Moore (1957) with adults.

Some of the investigators who have reported differences along lines that this study has failed to support have worked with persons far more deviant in regard to weight than those utilized in the present research. Perhaps such persons do exhibit characteristically different psychological traits, but these findings apparently cannot be extended to those moderately overweight as has frequently been implied in the literature of this problem (Bruch: 1941, 1951; Brosin, 1953).

Summary

This study attempted to detect differences between underweight, overweight, and normal weight college women in a number of psychological variables. Both self-report measures and excerpts from projective devices were used. The weight groups differed only in regard to subjectivity and social ascendency. It is suggested that these relationships may be accounted for by simpler explanations than have been thus far offered by others. The remainder of the variables under consideration revealed no significant departures from chance variations.

REFERENCES

Bett, W. R. Obesity: A survey of the recent literature. *Irish J. Med. Sci.*, 1951, **6**, 183–187.

Brosin, H. W. The psychology of overeating. *N. Eng. J. Med.*, 1953, **248**, 974–975.

Bruch, Hilde. Obesity in childhood and personality development. *Amer. J. Orthopsychiat.*, 1941, **11**, 467–475.

Bruch, Hilde. *The importance of overweight.* New York: Norton, 1951.

Conrad, S. W. The psychologic implications of overeating. *Psychiat. Quart.*, 1954, **28**, 211–224.

Dublin, L. I., & Marks, H. H. *The build of women and its relation to their mortality.* New York: Assoc. Life Insur. Med. Directors Amer., 1938.

Iversen, T. Psychogenic obesity in children. Vol. I. *Acta Pediat.*, 1953, **42**, 8-19.

Iversen, T., Juel-Nielson, N., Quaade, F., Tolstrup, K., & Ostergaard, L. On psychogenic obesity in children. *Acta Pediat. Supp.*, 1951, **40**, 59–60.

Kretschmer, E. *Physique and character.* London: Kegan Paul, 1925.

Laplane, R., & Etienne, M. Child obesity; psychologic data; one hundred cases. *Proc. Roy. Soc. Med.*, 1953, **46**, 1058–1059.

McHale, Kathryn. Comparative psychology and hygiene of the overweight child. *Teachers Coll., Columbia Univer. Contr. to Educ.*, 1926, No. 221.

Rennie, T. A. C. Obesity as a manifestation of personality disturbance. *Dis. Nerv. Sys.*, 1940, **11**, 238–247.

Richardson, H. B. Obesity as a manifestation of neurosis. *Med. Clin. N. Amer.*, 1946, **30**, 1187–1202.

Seltzer, C. C. Bodily disproportion and dominant personality traits. *Psychosom. Med.*, 1946, **8**, 75–97.

Sheldon, W. H., & Stevens, S. S. *The varieties of temperament.* New York: Harper, 1942.

Young, C. M., Beresford, K., & Moore, N. S. Psychologic factors in weight control. *Amer. J. Clin. Nutrition,* 1957, **5**, 186–191.

(Received November 20, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

SOCIAL DESIRABILITY AND THE EXPRESSION OF AGGRESSION UNDER VARYING CONDITIONS OF FRUSTRATION[1]

JOEL ALLISON AND DAVID E. HUNT
Yale University

That some persons attribute to themselves socially desirable traits when responding to self-report personality tests has long been recognized. However, there has been little systematic research on the specific effects of the tendency to "put oneself in a favorable light." Edwards' (1957) recent development of an instrument for measuring this tendency—the Social Desirability Scale (SDS)—presents an opportunity for studying the effects of the social desirability factor more systematically.

Consisting of 39 self-report items, the SDS is aimed to measure "the tendency of subjects to give socially desirable responses to statements in self description" (Edwards, 1957, p. 31). Correlations between the SDS factor and several current personality scales are quite high (Edwards, 1957). Therefore, it would seem important to find out more about the operation of this SDS factor: What responses are affected? What cues activate this factor? The present studies were designed to investigate the effect of the SDS factor upon the response of expressing aggression under varying hypothetical conditions of frustration.

Operating on the assumption that frustration will generally produce aggression, two studies were designed which varied the conditions of frustration in order to study these effects. Expression of aggression was selected as the dependent variable in the present studies because the response of verbal aggression is assumed to be socially undesirable in our culture, at least on a general level. Also, from an empirical standpoint, a correlation of −.75 has been reported between SDS and the Cook Hostility Scale (Merrill & Heathers, 1956). The studies were designed to test this assumption and in so doing to evaluate the construct validity (Cronbach & Meehl, 1955) of the SDS in this respect. Assuming in advance that social desirability will be associated with the suppression of aggression under at least *some* conditions, the more specific purpose of the present studies was to investigate the effect of situational conditions upon the relationship between social desirability and the expression of aggression. If the effects of social desirability could be anchored on both the stimulus and response sides, it was felt that a better understanding of social desirability as a process, in contrast to an empirical factor, might be possible.

Study I

Earlier investigators have demonstrated that the arbitrariness or justification of frustration is an important determinant of aggression (Cohen, 1955; Pastore, 1952), i.e., arbitrary or unjustified frustration is more likely to produce aggression than justified frustration. Therefore, the justifiability factor was varied as will be described.

Method

Subjects

Seventy male undergraduates were used as *S*s in developing the measure and 62 student nurses in the final procedure.

---

[1] The work described in this paper was supported by a research grant M-955 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. We wish to thank Theodore Anderson of Yale University and Theodore Lenn of New Britain (Conn.) State Teachers College for their assistance in providing subjects. We also wish to thank Norman Anderson of Yale University for his helpful advice on the statistical analysis.

Development of the Situational Frustration Test

A paper-and-pencil Situational Frustration Test was used to present the frustration variables. Twelve frustrating situations, each involving a different source of frustration, were presented to *S*s in this phase. These hypothetical situations will be referred to as unspecified, because the intention of the source or justification of frustration is not specified. An example of such an unspecified situation is: "You want to hold a party at your house. Your mother says that it will not be possible for you to use the house."

In order to specify these situations in terms of the justification factor, the 12 frustrating situations (in their unspecified form) were presented to 70 undergraduates with instructions to specify the conditions under which each of the sources of frustration would be considered justified or unjustified. An example follows:

> You want to hold a party at your house. Your mother says that it will not be possible for you to use the house.
>
> (*a*) Under what condition would you consider your mother's action unjustifiable or unwarranted? Write the first condition that occurs to you.
>
> (*b*) Under what condition would you consider your mother's action justifiable? Write the first condition that occurs to you.

These free responses were then analyzed in order to determine the conditions which would define each situation in its justified and unjustified form. These conditions served to define each of the 12 frustrating situations in a justified and unjustified form. For example, the situation already mentioned, in which the mother is the source of frustration, appeared in the following three forms:

> *Unspecified:* (See example above.)
>
> *Justified:* You want to hold a party at your house. Your mother says that it will not be possible for you to use the house because she has planned a party for her friends at the house at the same time.
>
> *Unjustified:* You want to hold a party at your house. Your mother says that it will not be possible for you to use the house because it would be too noisy and you might mess up the house.

The Situational Frustration Test[2] therefore consisted of each of the 12 situations in its three forms (unspecified, justified, and unjustified), resulting in a total of 36 items. The 12 unspecified items appeared first followed by the justified and unjustified items occurring in a randomized order. Following each of the 36 situations, the following question was asked:

> What would you do? [Circle one of the five choices in each row.]
>
> Get mad at. . . .
>
> (*a*) certainly would; (*b*) probably would; (*c*) might; (*d*) probably would not; (*e*) certainly would not

The statement of getting mad was made specific to each of the sources of frustration in each situation, e.g., "Get mad at my mother." The order of the five choices (*a* through *e*) was counterbalanced in each of the 36 items. These five choices were assigned values from 1 for "certainly would not" to 5 for "certainly would" in order to permit the calculation of scores for the expression of aggression. Therefore, for each *S* a score ranging from 12 to 60 could be determined for the expression of aggression under each of the three conditions: (*a*) unspecified frustration, (*b*) justified frustration, and (*c*) unjustified frustration. In addition, a total aggression score was available for each *S* by summing his scores in the three conditions.

In order to disguise partially the emphasis on aggression, two other response possibilities were included following each hypothetical frustration—"Would you keep your feelings to yourself?" and "Would you forget about it?"—with likelihood of response recorded by *S*'s selection of one out of five alternatives as before.

Procedure

The specific procedure was as follows: *S*s were 62 student nurses who were tested in one group. Each *S* first completed the Edwards 39-item SDS followed by the 36-item Situational Frustration Test.

The corrected split-half reliability of the

---

[2] Copies of both Situational Frustration Tests have been deposited with the American Documentation Institute. Order Document No. 6016, remitting $1.75 for microfilm or $2.50 for photocopies.

Table 1

Aggression Scores of Extreme SD Groups Under Three Conditions of Justification of Aggression

Condition	Low SD ($N=14$)	High SD ($N=14$)	Diff.	t	p
Unspecified	3.07[a]	2.56	.49	3.42	$<.01$
Justified	1.64	1.46	.18	1.22	$>.20$
Unjustified	3.54	3.36	.18	.89	$>.30$

[a] Higher scores indicate stronger expression of aggression.

total aggression score was .77. The correlation between the total aggression score and SDS and the correlation between the aggression score under each of the three conditions and SDS score were computed in order to determine the relationships between social desirability and expression of aggression. Specific situational effects were tested by comparing the aggression scores of two extreme groups on the SDS. Using defining scores of one standard deviation above and below the mean, a high and low group, each consisting of 14 *S*s, were selected.

Results

The correlation between SDS and total aggression scores was $-.27$ ($p < .05$)[3] or $-.34$ when corrected for attenuation; between SDS and expression of aggression in the unjustified condition $= -.14$; between SDS and expression of aggression in the justified condition $= -.18$; and between SDS and expression of aggression in the unspecified condition $= -.34$ ($p < .01$). The correlational analysis was also supported in the extreme group analysis shown in Table 1 in that the mean aggression score of the high SD group was significantly lower ($p < .01$) than the low SD group under the unspecified condition. However, when the incidence of aggressive responses under justifiable and unjustifiable conditions was considered, neither of the differences between the SD groups approached significance. Moreover, the difference between SD groups in the unspecified condition was significantly greater than the difference between SD groups in the justified condition ($p < .01$) and in the unjustified condition ($p < .05$).

Results in Table 1 are in agreement with earlier studies (Cohen, 1955; Pastore, 1952) that unjustified frustration is more likely to produce aggression than justified frustration. Further, this greater incidence of aggression under unjustified conditions may be noted in both the low SD ($t = 17.32$, $p < .001$) and high SD ($t = 15.09$, $p < .001$) groups.

These results suggest that the effect of the SD factor is strongest in a less structured, i.e., unspecified, frustrating situation. As more information becomes available, whether in the form of justification or nonjustification, the SD effect is no longer present.

Study II

Of the many factors besides justifiability which determine reactions to frustration (Dollard, Miller, Doob, Mowrer, & Sears, 1939), we selected two others in order to pursue further the possible situational specificity of the effect of social desirability: (*a*) strength of motive being frustrated and (*b*) availability of responses alternative to the frustrated response (whether or not an alternative means is present for circumventing frustration). That is, expression of aggression should generally increase as a function of increased motivation and decreased alternative responses. It should be noted that we are not presently concerned with studying these main effects of the motivation and alternative response condition. We wish to investigate the nature of SD effects upon expression of aggression throughout this range of conditions.

Method

Subjects

One hundred and seventy-one undergraduates, comprised of 82 males and 89 females, were used as *S*s.

Development of the Second Form of the Situational Frustration Test

Another form of the Situational Frustration Test (see Footnote 2) was developed for this study. Two variations of both motive strength and alternative response generated four classes of frustrating situations. Motivation was varied in the first portion of the description of the situation, while availability of response alternative was varied in the last portion of the description as the following examples indicate.

---

[3] All p values are based on two-tailed tests.

(*a*) *Low motivation—alternative response available.*

Someone has suggested to you that you have a party at your house but you are not sure whether or not you want to have one. Your mother says that it will not be possible for you to use the house. You think that you can probably use the apartment of a friend.

(*b*) *Low motivation—alternative response unavailable.*

Someone has suggested to you that you have a party at your house but you are not sure whether or not you want to have one. Your mother says that it will not be possible for you to use the house.

(*c*) *High motivation—alternative response available.*

You have been looking forward to having a graduation party for your closest friend. You set up all the plans. Your mother says that it will not be possible for you to use the house. You think that you can probably use the apartment of a friend.

(*d*) *High motivation—alternative response unavailable.*

You have been looking forward to having a graduation party for your closest friend. You set up all the plans. Your mother says that it will not be possible for you to use the house.

These examples present one of the basic situations expressed in each of the four forms. The test consisted of 12 such basic situations, each expressed in four forms, or a total of 48 items. Items were presented in blocks of 12 according to the above order, i.e., (*a*) first and (*d*) last, with items randomized within blocks. This order by blocks constituted a progressive increase in frustration which was deemed necessary in order to avoid possible "transfer effects" of strong frustrating situations preceding the weaker frustrating situations. In addition, analysis of responses to Study I as well as preliminary data indicated that there are no systematic differences in expression of aggression attributable to item position alone.

In Study II, *S*s again indicated the likelihood of their expressing aggression on a scale with five response alternatives which permitted determining aggression scores for the four conditions already described. Therefore, scores for each *S* were available for each of the four conditions and for the total of the four conditions. *S*s were tested in groups of 25 or 30 with the administration of the SDS preceding the situational frustration test. Extreme group analysis in this study was based on 25 *S*s in the low SD group (14 females and 11 males) and 23 *S*s in the high SD group (9 females and 14 males), extreme groups again defined by cutting scores of one standard deviation above and below the mean.

Table 2

Effect of Social Desirability and Nature of Frustration Upon Expression of Aggression

	df	*MS*	*F*	*P*
Social Desirability (High SD vs. Low SD)	1	2,887.04	19.98	<.001
Error	46	144.50		
Frustration	3	1,537.60	83.66	<.001
Social Desirability × Frustration	3	45.69	2.49	<.10
Error	138	18.39		
Total	191			

Results

The correlation between SD and total aggression for the 171 *S*s was .43 ($p < .01$). For the 82 males, r was .44 ($p < .01$), and for the 89 females, r was .42 ($p < .01$). Analysis of variance presented in Table 2 indicates that the aggression scores of the high SD group were significantly lower than those

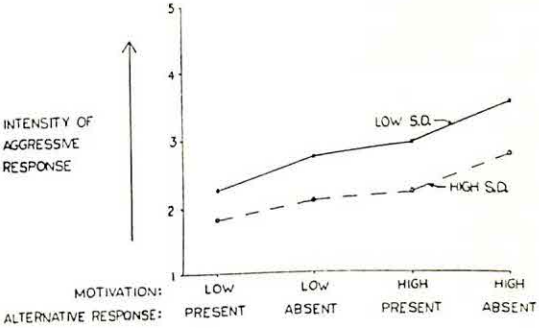


Fig. 1. Mean aggression scores of extreme groups in four frustration conditions.

of the low SD group ($p < .001$). In addition, there was a significant difference among the four conditions of frustrations ($p < .001$). Figure 1 presents these results by condition. When extreme group comparisons are made by condition, the low SD group expresses significantly more aggression in each of the four conditions of frustration. Extreme group comparisons were also made separately for males and females. In this analysis, seven of the eight comparisons led to significantly greater expression of aggression in the low SD group. Only the difference between extreme group males in the low motivation-alternative response present condition failed to reach significance at $< .05$ level.

Discussion

The findings of these two studies indicate that the social desirability factor exerts a situationally specific effect upon the verbal expression of aggression. In considering the results of Study I, it may be noted that the two conditions which specified the justification of the source of frustration also implicitly define the culturally approved response more clearly than in the unspecified condition. That is, aggression is usually sanctioned as a response to unjustified frustration and prohibited as a response to justified frustration. If we assume that the high SD group is sensitized to the cultural acceptability of a response, then we may note that under conditions of cultural acceptance, as implied in the unjustified frustration condition, the high SD group is capable of responding with verbal aggression. Furthermore, it is important to note that the effect of the SD factor is most pronounced when the situation is more ambiguous as in the unspecified condition.

That the high SD group expresses less aggression in all four conditions of Study II further emphasizes that the "suppressing" effect of the SD factor occurs primarily in situations in which the culturally acceptable response is not evident. That is, the information provided by the conditions in Study II pertained to the psychological or "internal" state of the recipient of frustration rather than to the intention of the source. There is a suggestion from the results of these two studies that persons high on the SD factor are, in Riesman's (1950) terms, "other-directed" rather than "inner-directed," but such a possibility requires further research support.

Summary

The relationship between Edwards' Social Desirability Scale (SDS) and expression of aggression as measured on a paper-and-pencil situational frustration test was investigated in two studies with the following results: When responding to frustrating situations in which the intention of the frustrating source was not specified, *S*s high on the SDS express significantly less aggression than *S*s low on the SDS. This greater tendency to express aggression in low SD *S*s was noted in four hypothetical variations involving the psychological state of the recipient of frustration. However, when socially normative information in the form of specifying the intention of the source was provided, the effect of the SD factor was no longer present.

The combined results of the two studies are taken to indicate construct validity for the Edwards' SDS and to emphasize the situationally specific role of the social desirability factor in determining the expression of aggression.

REFERENCES

Cohen, A. R. Social norms, arbitrariness of frustration, and status of the agent of frustration in the frustration-aggression hypothesis. *J. abnorm. soc. Psychol.*, 1955, **51**, 222–226.

Cronbach, L. J., & Meehl, P. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281–302.

Dollard, J., Miller, N. E., Doob, L. W., Mowrer, O. H., & Sears, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

Edwards, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

Merrill, R. M., & Heathers, Louise B. The relation of the MMPI to the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1956, **20**, 310–314.

Pastore, N. The role of arbitrariness in the frustration-aggression hypothesis. *J. abnorm. soc. Psychol.*, 1952, **47**, 728–731.

Riesman, D. *The lonely crowd.* Garden City, N. Y.: Doubleday, 1950.

(Received November 24, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

CORRELATES OF ATTITUDES AMONG CLINICAL PSYCHOLOGY STUDENTS[1]

SHELDON R. ROEN
University of New Hampshire

Distinctions have often been made between psychologists who freely draw interpretations directly from the material with which they work and those who prefer some quantitatively reliable instrument on which to base their conclusions. Shaffer (1953) investigated this distinction with a scale designed to measure "objective–intuitive" attitudes. In sampling members of Division 12, he found that objective and intuitive attitudes were highly related to the training the psychologists received and to the work they have done. A further question could be asked: does a student enter the clinical program with a preconceived set based on broader patterns in his personality, or is his subsequent orientation primarily a function of his perception of the expectancies of the significant person in his training? Conceptualizing the problem in this way focuses on the points of difference between role theory and the self-consistency hypothesis.

Role theory, according to Parsons and Shils (1952),

> . . . involves a set of complementary expectations concerning (an individual's) own actions and those of others with whom he interacts. . . . Roles are institutionalized when they are fully congruous with the prevailing culture patterns and are organized around expectations of conformity with morally sanctioned patterns of value orientations shared by members of the collectivity in which the role functions (p. 23).

In its most elementary form, role theory would be inclined to look for the current interpersonal setting surrounding the initial development of attitudinal patterns.

Self-consistency theory, on the other hand, emphasizes the prime need of an organism to maintain its mental organization as a unified whole. In the words of Lecky (1945, p. 81), "It is our view that behavior is usually 'in character' not because the separate acts are related to one another, but because all the acts of an individual have the goal of maintaining the same structure of values." Variations of the self-consistency hypothesis might include Adler's (1946) "style of life" construct and Goldstein's (1939) postulation of the one drive of "self-actualization." In its extreme form, this view conceives of individuality as a central, unitary, indivisible, striving process, and would look to the "basic nature" of an individual in understanding the reasons for his incorporation of new attitudes.

Method

A seven-page questionnaire was developed containing material in the following order: questions relating to biographical data; a shortened form of Eysenck's (1947, 1953) radical–conservatism and tough–tender minded factors; the restraint and masculinity scales from the Guilford-Zimmerman Temperament Survey (1949); 10 items each from Thorndike's (1954, 1955) helping individuals, psychometric (measurement), and experimenting activity preference clusters arranged as paired comparisons; multiple-choice questions relating to the profession and work of the most influential and significant person in the student's training; Shaffer's (1953) list of 21 adjectives from which a student could circle the attributes he felt were possessed by the most influential person; a multiple-choice question relating to the reasons for the student originally choosing to associate with this person; and Shaffer's (1953) objective–intuitive scale.

The psychology departments of 11 universities

---

[1] This paper, presented at the American Psychological Association meeting at Chicago, September 1956, is based upon a dissertation submitted to Teachers College, Columbia University, in partial fulfillment of the requirements for the degree of doctor of philosophy; the committee members were Laurance F. Shaffer, Edward J. Shoben, Jr., and Robert L. Thorndike. The author was a participant in the Army Senior Psychology Program at the time.

Table 1

Intercorrelations among Scaled Variables

Scale	T	R	Restraint	Masc.	Exp.–Help.	Meas.–Help.	Exp.–Meas.
Objective–intuitive	.03	.05	.01	.15**	.45**	.21**	.47**
Tough–tender minded (T)		−.34**	−.11	.08	.03	−.05	.04
Radicalism–conservatism (R)			−.13*	.03	.00	−.02	.10
Restraint				−.09	−.09	−.14*	−.04
Masculinity					.22**	.05	.25**
Experimenting–helping						.45**	.61**
Measurement–helping							.18**

* Significant beyond the .05 probability level.
** Significant beyond the .01 probability level.

were asked to cooperate by requesting their clinical psychology students to fill out the seven-page blank anonymously. Nine universities responded in time to be included in the main investigation, and 291 blanks were available for analysis. The sample represented 65% of the clinical students at the nine schools. The remaining two schools responded late and the data obtained from them was used only in comparing universities. Two of the universities were from the East Coast, five from the Midwest, and two from the West Coast.

Reliabilities of the eight scales included in the questionnaire were based on a random sample of 100 protocols. The split-half method, corrected by the Spearman-Brown formula, produced the following reliabilities: objective–intuitive scale, .86; helping, psychometric, and experimenting activity preferences, .89, .79, and .86 respectively; restraint, .67; masculinity, .77; tough–tender mindedness, .52; and radical–conservatism, .67. The homogeneity of the group sampled may have been responsible for lowering some of the coefficients.

Results

Relations among Scaled Variables

Table 1 presents the product-moment intercorrelations for the scaled variables. The top row contains the figures that bear more directly on the issues at hand. Neither of Eysenck's social attitudes seem related to the objective–intuitive dimension. The temperament trait of masculinity correlates beyond the .01 level of probability with the objective end of psychological orientation, as do all three of the activity preferences arranged as dual dimensions. A scatterplot of the activity preferences exhibited a skewness. Some students selected all the helping preferences and rejected all the experimenting activities. When scored as single dimensions, the helping, measurement, and experimenting preferences correlated −.46, −.27, and .53, respectively, with objective–intuitive attitudes. Each of these coefficients is significant beyond the .01 level of probability.

An inspection of the remainder of the table reveals other interesting relationships. Restraint appears to be slightly correlated with conservatism and the helping end of the measurement-helping dimension of preferences. The relationship between masculinity and objectivity is cross-validated, as it were, by its relationship to the experimental end of the experimental-helping and the experimental-measurement activity preferences.

Years of Training

An analysis was made of objective–intuitive attitudes with regard to year level in the clinical program. Mean scores and standard deviations were calculated for each educational level, and the distributions were further divided at the 30th and 70th percentile to see if strong attitudes were influenced as the student progressed in the program. An analysis of variance and *t* tests between means, standard deviations, and the differences between strong attitude means for year levels proved to be insignificant. Objective–intuitive attitudes would seem not to be a function of educational level.

Most Influential Person

Two multiple-choice questions asked the respondents to indicate the profession and work of the most influential and significant person in their training as psychologists. Of the categories that might have been selected as the profession of this person, only three (clinical psychologist, other psychologist, and psychiatrist) were chosen by enough students to be adequately compared. Table 2 summarizes the responses to the two questions. The only comparison not significant beyond the .05 level was that of psychiatrists vs. all others. However, when a test was drawn combining the psychiatrist and the clinical psychology groups as against all the others, a *t* even larger than that for the clinical psychology group alone was obtained. It would seem that students who were influenced by clinical psychologists or psychiatrists were significantly more intuitive than those who were influenced by other psychologists.

Table 2

Tests for Significance between Mean Objective-Intuitive Scores for Profession and Work of the Influential Person

Profession	Number	Mean	*SD*	*t*	*p*
Clinical psychologist	187	42.98	6.38	2.25	.012
All others	92	45.00	7.32		
Other psychologist	62	46.24	7.58	3.18	.001
All others	217	42.90	6.32		
Psychiatrist	25	42.02	6.30	1.34	.090
All others	254	43.80	6.79		
Psychiatrist & clin. psychologist	212	42.88	6.38	3.24	.001
All others	67	46.12	7.36		
Work					
Practice	109	41.67	5.82	4.28	.001
Research & Training	177	44.96	7.02		
Research	43	46.68	7.95	2.71	.003
Practice & Teaching	243	43.18	6.42		
Teaching	134	44.40	6.61	1.65	.049
Practice & Research	152	43.08	6.85		

The second section of the table suggests that students whose influential figures were engaged in practice were significantly more intuitive in their attitudes than those who were influenced by people primarily engaged in research. Students influenced by teachers elicited intermediate attitudes, but tended toward objectivity.

The adjectives circled as descriptive of the most influential person were conceived as falling into two groups, "warm" and "cool." Sympathetic, natural, and sensitive were included in the warm group, and hard headed, painstaking, and reserved formed part of the cool group. The congruence of these categories with perceptions of objectivity or intuitiveness is discussed by Shaffer (1953). The mean numbers of warm and cool adjectives circled by students scoring in the upper and lower 27% of the objective–intuitive distribution were compared and *t* ratios calculated. A similar analysis was made of adjectives previously found by Shaffer to be significant in distinguishing psychological attitudes. In each case the difference between the means was found to be significant beyond the .001 level of probability, indicating that objective–intuitive attitudes of students correspond to the descriptions they give of the influential person in their training.

Of the six items in the multiple-choice question that asked why a student associated with the person who influenced him most during his training, three of them (assigned to me; by chance or requirement; selected by him) were considered indicative of chance choices, and the other three (I selected him; I chose to interact with him; he impressed me most) were considered to reflect personal need choices. The space for a freely written answer, and the circling of both need and chance items by some of the respondents, left considerable latitude for scoring. Fourteen students wrote that their therapist was the most influential person. Several chi square computations were made between the number of students with strong and weak attitudes who indicated chance reasons, and those with strong and weak attitudes who gave need reasons. The upper and lower 27% of scores on the objective–intuitive scale defined the strong attitude group, and the middle 46% were considered to have weak attitudes. Significance was obtained when it was assumed that where both need and chance items were selected, need was prepotent, and where therapist was written in, this, too, reflected a need choice. Granting these conditions, it can be said with reservation that students with strong objective or intuitive attitudes selected their influential figures more on the basis of their own needs than for other considerations.

Comparison of Universities

The mean objective–intuitive score for the entire sample was 43.32, with a standard deviation of 6.80. When scores were grouped according to university, the means ranged from 40.03 for the school with the most intuitive students to 53.18 for the school with the most objective students. Objective–intuitive scores for graduates of the universities sampled were obtained from unpublished data gathered by Shaffer in his investigation. Ten or more scores of graduates were available for comparison for eight of the universities and a rank order correlation of .76 emerged between graduates and students. This is significant at the .03 level of probability.

Other Relationships

An analysis of objective–intuitive scores in relation to other data elicited by the questionnaire was made. Sex, internship status, and need or chance reasons for selecting the influential person (disregarding strong or weak attitudes) were not found to be meaningful. A highly significant difference in the intuitive direction was obtained between the means of students who indicated that a therapist was their most influential person and all the others. However, since the response was volunteered rather than solicited, and only 14 students responded in this way, little can be made of the finding.

The impressive results obtained in the comparisons of the profession and work of the most influential person suggested that relationships might exist between this and other variables as well. Professions and work were dichotomized on the basis of their relationship to objective–intuitive attitudes. Point-biserial correlations of .25 and .22, significant beyond the .01 level, were obtained between restraint and the profession and work, respectively, of the most influential person. No significant relationships with masculinity seemed to exist.

Discussion

Some of the evidence suggests that objective–intuitive attitudes of clinical psychology students are shaped by the training they receive, possibly in terms of learning roles. Students with intuitive attitudes indicated that the most influential persons in their training were "warm" or "intuitive" clinical psychologists or psychiatrists who were primarily engaged in practice. Objective students were influenced by other psychologists, described as "cool" or "objective," who engaged in research or teaching. The significant relationship in attitudes obtained between students and graduates of the same universities also tends to confirm this hypothesis.

However, the existence of a set or "self-consistency" motive among students prior to their entrance into a clinical program can also be supported by the data. Significant relationships were found to exist between an ingrained trait like masculinity and both objective attitudes and experimenting activity preferences. Restraint was associated with the profession and work of the influential figures. With some procedural reservations it was further found that students with strong attitudes, either objective or intuitive, chose to associate with the person that influenced them most more on the basis of their own needs than for other considerations. If experiences during graduate training were a primary factor, psychological orientation should become more distinct as a student advances in the program. This was not found to be true. Finally, it might also be argued that the strong relationships found to exist between objective–intuitive attitudes and the selection of future professional activities could support a predisposition hypothesis, since the type of work a student wants to do is usually decided before he enters a graduate training program.

It seems plausible that there are two kinds of student. One type could enter the program with a set toward a specific orientation and another could enter with an open mind, and thereby be more susceptible to influence. Other interpretations of the data might include the possibility that some students were influenced while undergraduates and are actually reflecting their relatively more immediate experiences than their more basic nature; or that students' attitudes develop very quickly during their first year of graduate training and continue for the most part as they were.

Summary

Objective–intuitive attitudes of clinical psychology students were investigated in an attempt to delineate the forces that might be operative in their development. Students from 11 universities responded to a blank containing measures of social attitudes, temperament traits, professional activity preferences, and objective–intuitive attitudes. Information about the most influential person in their training and personal biography was also solicited.

The data seemed amenable to various interpretations, including the possibility that many clinical psychology students entered a training program with a preconceived orientation toward objective–intuitive issues based, perhaps, on broader self-consistency patterns in their personalities. Other interpretations suggested were susceptibility to undergraduate or very early graduate influence, and that some students were self-consistent and others were role players.

REFERENCES

Adler, A. *Understanding human nature.* New York: Greenberg, 1946.

Eysenck, H. J. Primary social attitudes. I. The organization and measurement of social attitudes. *Int. J. Opin. Attitude Res.,* 1947, **1**, 49–84.

Eysenck, H. J. *The structure of human personality.* New York: Wiley, 1953.

Goldstein, K. *The organism.* New York: American Book, 1939.

Guilford, J. P., & Zimmerman, W. S. *The Guilford-Zimmerman temperament survey, manual of instruction and interpretations.* Los Angeles, Calif.: Sheridan Supply, 1949.

Lecky, P. *Self consistency: A theory of personality.* New York: Island Press, 1945.

Parsons, T., & Shils, E. A. (Eds.) *Toward a general theory of action.* Cambridge, Mass.: Harvard Univer. Press, 1952.

Shaffer, L. F. Of whose reality I cannot doubt. *Amer. Psychologist,* 1953, **8**, 608–623.

Thorndike, R. L. The psychological value systems of psychologists. *Amer. Psychologist,* 1954, **9**, 787–789.

Thorndike, R. L. The structure of preferences for psychological activities among psychologists. *Amer. Psychologist,* 1955, **10**, 205–207.

(Received December 8, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

REQUISITE CONDITIONS FOR BASIC PERSONALITY CHANGE[1]

ALBERT ELLIS

New York City

Many proponents of different schools of psychotherapy—such as Freud (1950), Rank (1945), Reich (1949), Reik (1948), Rogers (1957), and Sullivan (1954)—have claimed that certain conditions must exist and continue over a period of time if basic personality change is to be effected in disturbed individuals. It is the contention of this paper that there are probably no absolutely necessary conditions for constructive personality change.

Perhaps the clearest and most operationally stated set of requisites for change in the personality structure of the individual, at both surface and deeper levels, is that of Rogers (1957), who lists six "necessary and sufficient conditions": 1. Two persons (client and therapist) are in psychological contact. 2. The client is in a state of incongruence or anxiety. 3. The therapist is congruent or integrated in the relationship. 4. He experiences unconditional positive regard for the client. 5. The therapist experiences an empathic understanding of the client's internal frame of reference and endeavors to communicate this experience to the client. 6. The communication to the client of the therapist's empathic understanding and unconditional positive regard is to a minimal degree achieved. Although an initial consideration of Rogers' conditions makes them appear to be most reasonable and valid, more sober reflection provides some serious reasons to doubt their "necessity." Among these reasons are the following:

1. The necessity of two persons being in psychological contact for one of them to change his personality structure is seemingly belied by the thousands, perhaps millions, of significant personality changes that have occurred when a single individual (*a*) encountered external experiences and learned sufficiently by them to restructure his philosophy and behavior patterns of living or (*b*) without entering any real relationship with another, heard a lecture, read a book, or listened to a sermon that helped him make basic changes in his behavior.

2. The assumption that an individual must be in a state of incongruence or anxiety for personality change to occur is seemingly contradicted by the reasonably frequent instances of distinctly congruent and unanxious individuals who improve their personalities significantly by life experiences or reading or (as in the case of psychologists, psychiatrists, or social workers) benefit considerably from therapy which they entered largely for training purposes.

3. Although it is doubtlessly most desirable for a therapist who is trying to help a client to be, within the confines of their relationship, a congruent, genuine, integrated person, observation of a good many analysands or counselees may well lead the honest investigator to admit that some, though probably not many, have been considerably helped by their relationship with therapists who were (*a*) disturbed and incongruent individuals in their own right and, in some instances, (*b*) incongruent, hypocritical, and unintegrated in their relationship with these clients.

4. Although, again, the therapist's experiencing unconditional positive regard for his client is surely a desideratum of effective therapy, there seem to be more than a few cases on record where therapists have mainly

---

[1] Condensed version of a paper read at the workshop on psychotherapy of the American Academy of Psychotherapists, held in Madison, Wisconsin, August 9, 1958.

tried to regulate the lives and philosophies of their clients for the satisfaction of the therapists' own desires and where, nonetheless, emotionally exploited clients were considerably helped.

5. Even the therapist's experiencing an accurate, empathic understanding of the client's awareness of his own experience would appear to be desirable but unnecessary for effecting constructive personality change. More than one therapist has, with little or no emphatic understanding of his patients, indoctrinated some of them with a philosophy of life that proved to be considerably helpful; and some of the greatest bigots of all time, such as Savonarola, Rasputin, and Hitler, who, because of their own severe emotional disturbances almost certainly had a minimum of empathy with their fellow human beings, induced profound personality changes (at least a few of which were probably constructive) in their adherents.

6. It would seem, at first blush, almost indisputable that unless, as Rogers contends, the client perceives, to a minimal degree, the acceptance and empathy of the therapist, he will not significantly change. But second thought presents an open-minded theorist with the fact that apparently some paranoid patients become significantly improved in the course of therapy and still will not admit that their therapists really understand them, while others may recover from their paranoid thinking and only *after* significant personality change has been effected perceive that their therapists were accepting and empathic.

In the light of the foregoing considerations, it may perhaps be legitimately hypothesized that *very few* individuals significantly restructure their personalities when Rogers' six conditions are unmet; but it is most dubious that *none* do. Similarly, it is equally dubious that no patients make fundamental constructive improvements unless, as Freud (1950) contends, they undergo and resolve a transference neurosis during therapy; or, as Rank (1945) insists, they first have a highly permissive and then a strictly limited relationship with the therapist; or, as Reich (1949) claims, they loosen their character armor by having it forcibly attacked by the therapist's psychological and physical uncoverings; or, as Reik (1948) notes, they are effectively listened to by the therapist's "third ear"; or, as Sullivan (1954) opines, they undergo an intensive analysis of the security operations they employ with the therapist and with significant others in their environment. All these suggested therapeutic techniques may be highly desirable; but where is the evidence that any of them is *necessary?*

Are there, then, *any* conditions that are absolutely necessary if basic personality change in an individual is to occur? Probably not, if we think in terms of any *particular* condition or set of conditions. This null hypothesis can perhaps most quickly be tested if any theorist takes his own special theory of therapy and, with ruthless honesty, asks himself: Could *no* human being basically change his behavior were he and his therapist to ignore the essence of my theory and practice?

To test this hypothesis, the present author considered the essentials of his own special theory of rational psychotherapy which he has been developing for the last several years (Ellis: 1957, 1958a, 1958b, 1958c). The first fundamental assumption of this theory is that for real or basic personality change to occur, the individual must somehow learn to recognize his irrational, unrealistic perceptions and thoughts and change these for more logical, more reasonable philosophies of life. In honestly questioning whether this is a *necessary* condition for personality change, however, the following doubts arise:

1. Some people seem to make significant changes in their personalities without concomitantly acquiring notably new philosophies of living. It could be said, of course, that they really, unconsciously, *do* acquire such new philosophies. But this would be difficult to prove objectively.

2. Some individuals appear to change for the better when environmental conditions are modified, even though they retain their old childish views. Thus, a person who irrationally hates himself because he is poor may hate himself much less if he inherits a fortune. It could be said that the security he receives from inheriting the money really *does* make him change his childish, irrational views. But again: there would be difficulty in objectively validating this contention.

3. The statement that for constructive personality change to occur the individual must somehow basically change his thinking or his value system may be basically true; but it is also somewhat tautological. For in making such a statement one is really saying that poor personality integration consists of an individual's having unrealistic, self-defeating ideological assumptions and that to change his personality integration for the better he must surrender or change these assumptions. Although descriptively meaningful, this statement boils down to the sentence: in order to change his personality the individual must change his personality; or: in order to get better he must get better. This proves very little about the "necessary" conditions for personality change.

Again: rational psychotherapy significantly differs from virtually all other theories and techniques in that, according to its precepts, it is desirable not merely for the therapist to uncover, understand, and accept the client's illogical and unrealistic assumptions which cause him to remain immature and ineffective, but it is usually required that he forthrightly and unequivocally *attack* and *invalidate* these assumptions. Is this desiderata of psychotherapy necessary? Most probably not: since *some* clients and nonclients (although relatively few, I believe) seem to have significantly improved in spite of their not having the benefit of a competent rational therapist to help them understand how they acquired, how they are currently sustaining, and how they can and should forthrightly attack and annihilate their basic irrational attitudes and assumptions.

The conclusion seems inescapable, therefore, that although basic constructive personality change—as opposed to temporary symptom removal—seems to require fundamental modifications in the ideologies and value systems of disturbed individuals, there is probably *no* single condition which is absolutely necessary for the inducement of such changed attitudes and behavior patterns. Many conditions, such as those listed by Freud, Rank, Reich, Reik, Rogers, Sullivan, and other outstanding theorists, or such as I list in my papers on rational psychotherapy, are highly desirable; but all that seems to be necessary is that the individual *somehow* come up against significant life experiences *or* learn about others' experiences *or* sit down and think for himself *or* enter a relationship with a therapist who is *preferably* congruent, accepting, empathic, rational, forceful, etc. Either/or rather than this-and-that seems to be the only realistic description of necessary conditions for basic personality change that can be made at the present time.

REFERENCES

Ellis, A. Outcome of employing three techniques of psychotherapy. *J. clin. Psychol.*, 1957, **13**, 344–350.

Ellis, A. Neurotic interaction between marital partners. *J. counsel. Psychol.*, 1958, **5**, 24–28. (a)

Ellis, A. Rational psychotherapy. *J. gen. Psychol.*, 1958, **59**, 35–49. (b)

Ellis, A. Hypnotherapy with borderline schizophrenics. *J. gen. Psychol.*, 1958, **59**, 245–253. (c)

Freud, S. *Collected papers.* (Originally published 1924.) London: Hogarth, 1950.

Rank, O. *Will therapy; and, Truth and reality.* New York: Knopf, 1945.

Reich, W. *Character analysis.* New York: Orgone Institute, 1949.

Reik, T. *Listening with the third ear.* New York: Farrar, Straus, 1948.

Rogers, C. The necessary and sufficient conditions for therapeutic personality change. *J. consult. Psychol.*, 1957, **21**, 95–101.

Sullivan, H. S. *The psychiatric interview.* New York: Norton, 1954.

(Received December 9, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

A NOTE ON AN INDEX FOR THE COMPARISON OF PROFILES ON THE "CIRCLE"[1]

WILLIAM L. HAYS
University of Michigan

An Index for Two Individuals

The problem of the comparison of profiles on the coding device known as the "Circle" (Freeman, Leary, Ossorio, & Coffey, 1950) arises whenever behavior of an individual has been coded in this way and one wishes to compare his pattern of behavioral trends with the pattern for another individual (or with the pattern for the same individual at another time). In the past this problem has been handled simply by doing a chi square comparison between the relative frequencies of codings in the various categories of the Circle for the profile. Another alternative has been to rank order the relative frequencies in the categories for the two individuals and to compute a rank correlation measure between them. Both of these procedures, however, have been considered inadequate since they seem to do violence to the essential idea of the Circle: this is, that the categories of the Circle are assumed to be ordered in the fashion reflected by their arrangement on the circle. Such techniques seem to ignore the "circularity" of the Circle categories when the pattern comparisons are made.

The technique proposed herein is designed to incorporate this idea of a circular ordering into a measure of pattern similarity between individuals. The unit of measurement explicitly adopted here is the degree of adjacency between two categories. Thus it assumes that two observations falling into the same category are, for all relevant purposes, identical. Two observations falling into adjacent categories are at a unit distance; two observations falling into categories separated by one intermediate category are at Distance 2; etc. The most dissimilar (distant) two observations coded into the Circle can be is then taken to be 8.

Using this assumption, we may then define the "scatter" of the observations coded for an individual into the circle to be the average distance between all pairs of observations in his circle pattern. Furthermore, the *dissimilarity* between two profiles (of Individuals I and II) may be taken to be the average distance on the circle between pairs of observations belonging to *different* individuals, minus the average distance on the circle between pairs of observations belonging to the *same* individual. This gives rise to a measure of *dissimilarity* between profiles which can be stated as follows: the "Circle-Dissimilarity" of Individuals I and II is given by

$$CD = \sum_i (X_{i\,I} - X_{i\,II}) \sum_j d_{i\,j}(X_{j\,II} - X_{j\,I})$$

where:

$X_{i\,I}$ = the proportion of observations for Individual I falling into Category i
$X_{i\,II}$ = the proportion of observations for Individual II falling into Category i
$X_{j\,I}$ = the proportion of Individual Is observations falling into Category j
$X_{j\,II}$ = the proportion of observations for Individual II falling into Category j
$d_{i\,j}$ = the *distance* on the Circle between Categories i & j (where $d_{i\,i} = d_{j\,j} = 0$)

This measure is very easily computed for a pair of individuals by means of the computing sheet (Tables 1 and 2). This includes a table with numbers in the body of the table (standing for the distances $d_{i\,j}$) and with letters

---

[1] The work reported here was carried out in connection with a research program on psychotherapy supported by grant M-516 C-3 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

THE SENTENCE COMPLETION TEST AS A PREDICTOR OF CONTINUATION IN PSYCHOTHERAPY[1]

E. WESLEY HILER

Agnews State Hospital, Agnew, California

One of the problems with which most mental hygiene clinics are faced is the premature termination of psychotherapy by patients. Previously we have reported on differences in the Rorschach and Wechsler performances between patients who continued and patients who discontinued therapy, and our efforts to predict which patients will terminate prematurely (Gibby, Stotsky, Hiler, & Miller, 1954; Hiler, 1958b). The present article will deal with the sentence completion test performance of these criterion groups.

Method

The present study was carried out at the Veterans Administration Mental Hygiene Clinic at Detroit, Michigan. This clinic is analytically oriented and tends to favor long-term methods of psychotherapy. For a more complete description of the patients, therapists, and clinic procedures, see a previous publication (Gibby et al., 1954).

The sentence completion test records of the same sample of patients used for the Rorschach study were used in the present study. In that study we had used as many records as were available in the files and had matched remainers and terminators with respect to therapist, i.e., we had the same number of remainers as terminators for each therapist. Unfortunately, only about one third of the sample had been administered the sentence completion test so our sample is much smaller and no longer perfectly matched with respect to therapist in the present study.

As in the Rorschach study, we used extreme groups for our comparisons. Patients who terminated within 5 sessions were considered to have terminated prematurely and were compared with patients who remained in treatment for 20 or more sessions. The particular sentence completion test used was the Michigan Sentence Completion Test. This test consists of 100 items and was developed in connection with the assessment of clinical psychologists (Coplin, 1951; Kelly & Fiske, 1951) and revised for use with psychiatric patients. It is structured with the specific purpose of obtaining information about the patient's strongest drives, his principal conflicts, his attitudes toward his parents, siblings, wife, employer, etc. Many items are designed to reveal the way the patient deals with his aggressive impulses, anxiety, and guilt as well as the intensity and sources of these feelings.

For the purpose of the present study, special scoring categories were developed for each item by examining the records of an exploratory sample of remainers and terminators ($N = 25$). Categories were chosen which seemed to reveal differences between the criterion groups on this sample.

A cross-validation sample ($N = 70$) was then scored on the basis of these response categories. The frequency of responses falling in each category was then tabulated, permitting a chi square comparison between the criterion groups for each item of the test.

In an effort to characterize the criterion groups in a more reliable fashion than is provided by individual item analysis, 25 personality dimensions were selected, each of which could be evaluated by a series of items of the sentence completion test. For instance, on the variable "desire for status" all items in which the subjects could express their longing for success, fame, or wealth were scored 0, +1, or +2 on this dimension depending on the apparent intensity of their expressed desire. The total score on "desire for status" would be the sum of the individual item scores on that dimension.

---

[1] This report is based on a dissertation submitted to the Department of Psychology at the University of Michigan and was carried out at the Veterans Administration Mental Hygiene Clinic in Detroit, Michigan.

The scoring criteria were developed on a logical or a priori basis with the aid of certain scales already in existence (Coplin, 1951; Forer, 1950; Rohde, 1946; Rotter & Willerman, 1947). The criteria used in scoring each item on each variable are fairly objective and may be found in the appendix of the doctoral dissertation on which this study is based (Hiler, 1953).

A three by two chi square analysis was used to compare the criterion groups on each of these variables. This method of statistical analysis was used since it reveals extremes vs. middle position differences as well as high vs. low differences between the criterion groups.

A prediction formula was developed using items with response categories which discriminated the best between the criterion groups on the exploratory sample. The formula (Hiler, 1953, pp. 161–162), which discriminated with 100% accuracy on the exploratory sample, was then cross-validated on our cross-validation sample.

Also, a clinician, experienced in the use of the sentence completion test, was given the total sample of 95 sentence completion protocols, without any other information, and was asked to identify the patients whom he thought would terminate prematurely on the basis of his clinical judgment.

Results and Discussion

Out of 100 sentence completion test items, 15 appeared to discriminate between the criterion groups beyond the .05 level of confidence as shown in Table 1.

The differences between the remainers and the terminators on these items present a fairly consistent picture. One of the main findings is the greater willingness to reveal personal feelings and confidential material shown by the patients who remained in treatment. This is shown most clearly in items dealing with attitudes toward parents and toward the self. The patients who subsequently remained in treatment are willing to go into detail on these topics, whereas the terminators often resort to evasive clichés and vague overgeneralization. For instance, the remainers often characterized both parents on Item 12, whereas the terminators frequently maintained that there was no difference between their parents.

Another characteristic of the remainers (related to the above) is their frank admission of feelings of inferiority and social inadequacies. Thus, on Item 17, the remainers report feeling bad about themselves and their personal inadequacies or shortcomings, whereas the terminators more often mention such things as unfavorable circumstances or "things in general." And on Item 22, the remainers mention social inadequacies as causing them the greatest amount of trouble in their youth, whereas the terminators more frequently mention acts of delinquency or situational factors.

A third characteristic of the patients who remained in treatment is the tremendous need for achievement, success, and status revealed in their sentence completion responses. This may in part represent an attempt to compensate for the inferiority feelings mentioned above, or it may represent a higher energy level and greater determination or drive which Barron (1953) has included under the general term "ego strength" and has reported to be related with improvement under psychotherapy. This would be consistent with our Rorschach findings in which the remainers showed greater productivity (Gibby et al., 1954).

A fourth characteristic of the remainers is their greater psychological sophistication. In general, they tend to have more insight into psychological causation than do the terminators. For instance, on 57, more of the remainers mentioned some form of *mental* cruelty as being the worst thing a woman can do to a man. Likewise on Item 67, some of the remainers mentioned *psychological* forms of cruelty, whereas none of the terminators included these, mentioning instead *physical* acts of violence. The remainers also showed greater awareness of emotional needs. For instance, on Item 34, most of the remainers mentioned some real emotional need, whereas many terminators mentioned such things as baseball,

Table 1

Percentage of Responses of Remainers and Terminators Falling in Various Categories on Sentence Completion Test Items

Item	Content	Percentage of Responses in Each Category		
		Remainers ($N = 41$)	Terminators ($N = 29$)	
4	People would do anything in order to . . .			
	(*a*) need for status, wealth, or success	.83	.40	
	(*b*) all other responses	.17	.60	$p = .001$*
5	I pity . . .			
	(*a*) people in general	.07	.27	
	(*b*) people with bad characteristics	.15	.14	
	(*c*) people with illness or handicaps	.22	.31	
	(*d*) people in unpleasant circumstances	.32	.07	
	(*e*) other responses	.24	.21	$p = .05$
12	The difference between Mom and Dad was . . .			
	(*a*) both characterized	.22	.03	
	(*b*) one characterized	.24	.31	
	(*c*) physical qualities (age, sex)	.22	.21	
	(*d*) no difference	.00	.28	
	(*e*) other	.32	.17	$p = .001$
14	What meant most to me . . .			
	(*a*) other people	.20	.35	
	(*b*) being liked or being happy	.24	.41	
	(*c*) other (specific things or events)	.56	.24	$p = .05$
17	I used to feel bad about . . .			
	(*a*) specified other people	.22	.10	
	(*b*) myself (personal inadequacies)	.49	.31	
	(*c*) other	.29	.59	$p = .05$
22	As a youth my greatest trouble was . . .			
	(*a*) aggression and delinquency	.05	.28	
	(*b*) social inadequacies	.29	.14	
	(*c*) other	.66	.59	$p = .02$
34	As a youth my greatest emotional need was . . .			
	(*a*) parental love implied	.32	.14	
	(*b*) nonemotional needs mentioned	.15	.38	
	(*c*) other	.54	.48	$p = .02$
37	A white man who has intercourse with a black woman . . .			
	(*a*) tolerant attitude expressed	.27	.48	
	(*b*) intolerant attitude expressed	.51	.52	
	(*c*) nonevaluative statements	.22	.00	$p = .02$

Table 1 (*Continued*)

Item	Content	Percentage of Responses in Each Category		
		Remainers ($N = 41$)	Terminators ($N = 29$)	
41	I like people at parties who . . .			
	(*a*) are fun-loving, good mixers	.49	.59	
	(*b*) friendly, kind, courteous	.29	.07	
	(*c*) well behaved, controlled	.07	.28	
	(*d*) other	.15	.07	$p = .02$
55	Love . . .			
	(*a*) concept described	.37	.07	
	(*b*) favorable attitude expressed	.27	.52	
	(*c*) other	.37	.41	$p = .02$
57	The dirtiest thing a woman can do to a man is . . .			
	(*a*) be unfaithful	.49	.62	
	(*b*) leave him	.15	.00	
	(*c*) mental cruelty	.29	.14	
	(*d*) other	.07	.24	$p = .02$
67	One must never . . .			
	(*a*) kill or injure people	.20	.07	
	(*b*) mental cruelty	.15	.00	
	(*c*) be mad, do wrong	.12	.35	
	(*d*) other	.54	.59	$p = .02$
72	The most significant thing about my school experience . . .			
	(*a*) favorable events	.24	.07	
	(*b*) unfavorable events	.20	.07	
	(*c*) neutral	.22	.48	
	(*d*) other	.34	.38	$p = .05$
97	People should tell lies when . . .			
	(*a*) they feel like it (mere volition)	.00	.17	
	(*b*) necessary or no harm done	.73	.38	
	(*c*) never	.27	.45	$p = .01$
100	Authority . . .			
	(*a*) only for those who know how to use it	.39	.10	
	(*b*) unfavorable attitude expressed	.17	.17	
	(*c*) favorable attitude expressed or other	.44	.72	$p = .05$

Note.—Percentages are included in this table to facilitate comparisons.
* The probability values are based on χ^2 computed from the actual frequencies.

travel, or education as the greatest emotional need in their youth.

The terminators, in addition to being in general more evasive, stereotyping, and psychologically unsophisticated, also contained individuals on the extremes of superego domination. Some terminators seem to have a rigid superego, which is very intolerant of impulse gratification as is manifested in their response to Item 41, where many terminators stress good behavior and self-control as being of primary importance at parties. In contrast to these are a group of terminators showing evidences of psychopathic tendencies. Thus on

97, a number of terminators maintained that people should tell lies *whenever they want to.* None of the remainers answered in this way. Also, many of the terminators seemed to have a history of delinquent behavior, as reflected in their answers to Item 22, where a number of terminators mentioned aggression or delinquency as the greatest problem of their youth.

For the results on the scales developed for the test as a whole, see Table 2.

Although almost all the differences between the criterion groups are in the direction one would expect, only 5 out of the 25 differences are great enough to achieve statistical significance beyond the .05 level of confidence. Willingness to reveal self on test, feelings of inadequacy, desire for status, and objectivity are associated with remaining in therapy; and evasiveness is associated with premature termination of therapy. These results are all consistent with our previous discussion on the individual item findings. Note that there is practically no difference between the remainers and terminator on items omitted and on our measures of dependency need, paranoid tendencies, and overt hostility.

Table 2

Comparison of Cross-Validation Sample Criterion Groups on Sentence Completion Variables

Variable	χ^2 (2 *df*)	Direction	*p*
Omissions	.35	none	NS
Evasiveness	12.73	–	.005
Positive façade	3.38	–	.20
Denials	2.30	–	NS
Positive orientation	4.21	–	.20
Negative orientation	3.93	TRT	.20
Desire for status	6.14	+	.05
Superego domination	4.05	–	.20
Pleasure domination	1.36	–	NS
Affiliation need	1.02	+	NS
Dependency need	.08	none	NS
Passivity in face of aggression	1.93	–	NS
Altruism	4.15	+	.20
Objectivity	7.20	+	.025
Depression	4.10	+	.20
Paranoid feelings	.26	none	NS
Overt hostility	.22	none	NS
Somatic preoccupation	3.93	–	.20
Feelings of inadequacy	12.26	+	.005
Motivation for recovery	3.77	+	.20
Complexity of response	2.99	+	NS
Willingness to reveal self	14.37	+	.001
Bizarreness	4.75	–	.10
Silliness	.51	–	NS
Reference to alcoholism	3.00	–	NS

Note.—+ Variable shows positive relation to continuance; – Variable shows negative relation to continuance; TRT Remainers occupy intermediate position; NS Not significant even at .20 level.

The prediction formula developed on the exploratory sample discriminated between the criterion groups with 71% accuracy on the cross-validation sample. This is approximately the same as our best Rorschach (68% accuracy) and Wechsler-Bellevue (65% accuracy) predictors (Gibby et al., 1954; Hiler, 1958b) and may be near the maximum degree of accuracy possible when the prediction is based on the records of patients alone without consideration of the therapist's personality and competence in making the prediction. It has been found that different therapists lose different types of patients (Hiler, 1958a).

The clinician, using his intuitive or global judgment rather than a prediction formula, achieved 68% accuracy of prediction. This figure is not much below that obtained by the sign approach and suggests that the sentence completion test, perhaps because of its high degree of face validity, can yield fairly accurate predictions even when the intuitive method of clinical judgment is used.

It is the opinion of the author that the sentence completion test is potentially one of the most useful devices for selecting patients for psychotherapy. It seems to have considerable face validity, for it permits fairly accurate predictions on the basis of intuitive or clinical judgment. This makes it especially suited for the many clinicians who prefer the use of clinical judgment over the actuarial sign approach. The sentence completion test seems to permit inferences about many aspects of the patient's personality which are related to his behavior in psychotherapy. Items could be especially designed to measure the variables found to be related to continuation in psychotherapy and other variables, which, on empirical or theoretical grounds, one would hypothesize to be related to the patient's ability to enter into and profit from a psychotherapeutic relationship.

Summary

Sentence completion responses of patients who remained in treatment for at least 20 sessions were compared with the responses of patients who terminated within 5 sessions.

Significant differences were found between these groups. Patients who subsequently remained in treatment tended to be:

1. Less evasive and more willing to reveal personal feelings and confidential material about themselves and their parents.
2. More preoccupied with feelings of personal and social inadequacy.
3. More driven by needs for achievement, success, and status.
4. More aware of emotional needs and in general showing greater psychological sophistication.

Patients who terminated included a number with overly rigid superegos as well as a number of individuals with rather weak superegos.

A prediction formula was developed on an exploratory sample which permitted 71% accuracy of prediction on a cross-validation sample.

A clinician using the intuitive method of clinical judgment predicted the criterion with 68% accuracy.

REFERENCES

Barron, F. An ego-strength scale which predicts response to psychotherapy. *J. consult. Psychol.*, 1953, **17**, 327–333.

Coplin, H. The measurement of subject's orientation toward a sentence completion test. Unpublished doctoral dissertation, Univer. of Michigan, 1951.

Forer, B. R. A structured sentence completion test. *J. proj. Tech.*, 1950, **14**, 15–30.

Gibby, R. G., Stotsky, B. A., Hiler, E. W., & Miller, D. R. Rorschach criteria for predicting duration of therapy. *J. consult. Psychol.*, 1954, **18**, 185–191.

Hiler, E. W. An investigation of psychological factors associated with premature termination of psychotherapy. Unpublished doctoral dissertation, Univer. of Michigan, 1953.

Hiler, E. W. An analysis of patient-therapist compatibility. *J. consult. Psychol.*, 1958, **22**, 341–347. (a)

Hiler, E. W. Wechsler-Bellevue Intelligence as a predictor of continuation in psychotherapy. *J. clin. Psychol.*, 1958, **14**, 192–194. (b)

Kelly, E. L., & Fiske, D. W. *The prediction of performance in clinical psychology.* Ann Arbor: Univer. Michigan Press, 1951.

Rohde, A. R. Explorations in personality by the sentence completion method. *J. appl. Psychol.*, 1946, **30**, 169–181.

Rotter, J. B., & Willerman, B. The incomplete sentences test. *J. consult. Psychol.*, 1947, **11**, 43–48.

(Received December 23, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

A FACTOR STUDY OF THE MMPI PSYCHOPATHIC DEVIATE SCALE

ALEXANDER W. ASTIN
United States Public Health Service Hospital, Lexington, Kentucky

The clinical diagnostic term "psychopathy" has been the target of extensive criticism during recent years. Its supposed ambiguity (Cameron & Magaret, 1951), among other things, has been one of the principal features which has stimulated comment. Indeed, the authors of the MMPI[1] Psychopathic Deviate (*Pd*) scale early recognized the empirical multidimensionality of the psychopath concept (McKinley & Hathaway, 1956).

Extended use by the author of the MMPI with a population of hospitalized drug addicts led to the impression that equally high *Pd* scores can be obtained by vastly different types of "psychopaths." The present study grew out of these experiences and was concerned primarily with two problems: (*a*) identifying, by means of a factor analysis of the MMPI *Pd* scale, some of the traits which characterize people labeled psychopathic deviates; and (*b*) demonstrating that identical *Pd* scores can have quite different clinical implications, depending upon the composition of the internal factors contributing to the total *Pd* score.

Subjects

All *S*s were male narcotic addicts[2] from the U. S. Public Health Service Hospital, Lexington, Kentucky. Two hundred-fifty *S*s were selected at random from the hospital population for the factor analytic portion of the study. Eighty-eight additional *S*s, including 23 physician addicts, were used in the second part of the study. The age range of the *S*s was 19–61, with three fourths of the *S*s between ages 20 and 35. *S*s whose raw scores on the MMPI *F* scale surpassed 15 were discarded and replaced by *S*s with valid profiles. The mean *Pd T* score for male addicts at this hospital is 75, and approximately 60% obtain their highest MMPI score on this scale. Less than 15% obtain *Pd* scores ranking lower than third among the nine clinical scales. However, the variability of the addicts on this scale is comparable to that of normals (standard deviation = 10.6 *T* score points).

Method

The method used in the factor analytic portion of this study was identical to that used by O'Conner, Stefic, and Gresock (1957), who recently factored the MMPI Depression scale. Using the sample of 250 addicts with valid profiles, tetrachoric correlation coefficients were computed among the 50 *Pd* scale items. Items with response frequencies less than 10 or greater than 90% were discarded, reducing the *Pd* item total to 49. The resulting matrix of 1,176 intercorrelations was subjected to a multiple-group factor analysis. Rotation to oblique, simple structure was accomplished by the single plane method.

After the factors had been identified, two attempts were made to demonstrate that identical *Pd* scores can have different clinical implications. The first involved locating pairs of depressed and nondepressed *S*s who could be matched on both *Pd* and its *K* correction.[3] Depressed *S*s were defined as those having the MMPI *D* (Depression) score above $T = 75$; nondepressed *S*s were those with *D* below $T = 57$. A depressed and a nondepressed *S* were considered matched if the *Pd* and *K*, respectively, of one *S* were each

---

[1] Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1951).

[2] These *S*s were not currently addicted and were recovered from the debilitating effects of opiate withdrawal.

[3] Matching on *K* was also considered necessary, in that .4*K* is normally included as part of the total *Pd* score.

Table 1

Rotated Factor Matrix

Pd Items[a]	I	II	III	IV	V
1	−.30	.08	.20	.01	.23
2	−.35	.21	.00	−.11	.22
3	−.07	.31	−.15	.10	.35
4	−.45	.20	.05	−.38	.29
5	−.05	.69	−.11	−.08	.25
6	−.25	.26	.18	−.10	.12
7	−.32	.07	.02	−.71	.19
8					
9	.06	−.01	−.18	.01	.01
10	−.05	−.09	.73	−.05	.22
11	.05	.33	.33	.10	.04
12	−.41	.21	.17	−.09	.18
13	−.65	.21	.22	−.05	.38
14	.13	−.34	.17	.08	−.17
15	−.21	.19	.17	−.22	.05
16	−.28	.16	.06	−.05	.55
17	−.54	.28	.28	−.09	.30
18	−.23	.28	.09	−.15	.33
19	−.66	.08	.29	−.17	−.03
20	−.66	.55	−.09	−.10	.22
21	−.27	.26	.03	.22	.35
22	−.10	.68	.30	−.11	.06
23	−.02	.08	.72	−.14	.00
24	.05	.04	.31	−.07	−.36
25	.35	−.08	−.28	.36	−.10
26	−.20	.26	.13	.22	.63
27	.25	−.43	−.26	.06	−.02
28	−.07	−.11	.16	.08	.18
29	−.36	.06	.05	.07	.63
30	.62	−.05	−.02	.15	−.23
31	−.34	.29	.74	.17	.18
32	.70	−.26	−.05	.05	−.34
33	−.11	.03	.10	.35	−.02
34	.56	−.27	.09	.29	−.39
35	−.20	.09	.11	−.03	.03
36	−.37	.04	−.09	−.23	.68
37	−.26	.28	.41	−.21	−.06
38	.21	−.05	−.17	.57	−.08
39	−.24	.01	.05	−.31	−.03
40	.02	.20	.02	−.01	.13
41	−.34	.15	.01	−.26	.27
42	−.29	.39	.28	−.45	−.03
43	−.32	.19	.27	−.07	.04
44	.27	−.28	−.05	.50	.32
45	.66	−.44	−.01	.18	−.17
46	−.28	.59	.06	−.25	−.03
47	−.07	−.03	.08	.48	.17
48	.22	−.12	.25	.21	.08
49	−.07	−.14	.42	.02	−.09
50	.22	−.03	−.23	.67	.01

[a] Items are numbered in order of their appearance in the inventory.

within one raw score point of the *Pd* and *K* of the other *S*. Twenty-one such matched pairs were collected. Mean differences between depressives and nondepressives on each of the *Pd* factors were tested for significance by means of *t* tests for matched samples ($df = 20$).

The final comparison attempted to relate *Pd* factor differences to a non-MMPI variable. Twenty-three physician addicts were matched with 23 nonphysician addicts on *Pd* and *K* (as above). *t* tests for matched samples ($df = 22$) were used to test differences between physicians and nonphysicians on each of the *Pd* factors.

Results and Discussion

Five identifiable factors emerged from the analysis of *Pd* item intercorrelations.[4] Rotated loadings of the 49 items on each factor are presented in Table 1.

Sixty-two per cent of the 1,176 residuals fell between ±.09, inclusively, and 91% fell within ±.19. Since close inspection of the larger residuals did not reveal any meaningful clusters involving more than two items, no further factors were extracted. Several of the obtained communalities differed markedly from the initial estimates, as indicated by an *r* of only .58 between the two sets of estimates. The larger discrepancies tended to involve items with extreme response frequencies, where initial estimates of communalities (centroid method) often appeared to have been based on spuriously high intercorrelations.

Each of the five factors, together with items containing the highest loadings, will be presented and discussed separately.

Factor I: Self-Esteem

Item	Oblique Loading[5]
32. I find it hard to make talk when I meet new people. (F)	.70
45. When in a group of people I have trouble thinking of the right things to talk about. (F)	.66
20. Much of the time I feel as if I have done something wrong or evil. (T)	−.66

---

[4] Copies of the correlation matrix may be obtained from the American Documentation Institute. Order Document No. 6066, remitting $1.25 for microfilm or $1.25 for photocopies.

[5] When scored in the keyed (i.e., *Pd*) direction.

Item	Oblique Loading[5]
19. My hardest battles are with myself. (T)	−.66
13. I wish I could be as happy as others seem to be. (T)	−.65
30. It makes me uncomfortable to put on a stunt at a party even when others are doing the same sort of things. (F)	.62
34. I wish I were not so shy. (F)	.56
17. I do many things which I regret afterwards (I regret things more or more often than others seem to). (T)	−.54
4. At times I have very much wanted to leave home. (T)	−.45
12. I have not lived the right kind of life. (T)	−.41

The positive pole of this factor (Items 30, 32, 34, 45) consists of statements denying social introversion and shyness; the negative pole (Items 4, 12, 13, 17, 19, 20) contains statements admitting depression and guilt feelings. In some ways these poles resemble the two major varieties of social deviates cited by Cameron and Magaret (1951, pp. 198–206). These authors describe the "inadequate" (low Self-Esteem?) social deviant as one who fails "to develop and maintain effectively the role of social participant," one who partially recognizes his social inadequacies and therefore seeks a dependent, protective relationship with others. The "inept" (high Self-Esteem?) social deviant, on the other hand, has acquired techniques of social participation, but uses them abortively in his attempts to interact with the society.

That both ends of a bipolar factor should exist within the same clinical scale is somewhat surprising. This means that persons who respond to one pole in the *Pd* direction will tend *not* to respond to the other pole in the *Pd* direction. Psychologically this phenomenon makes sense, particularly in light of the interpretations of the two sets of items; however, the *psychometric* implications of such a finding are difficult to rationalize. That is, a scale designed to measure a given syndrome, however heterogeneous, is not expected to contain mutually exclusive subsets of items.

Factor II: Hypersensitivity

Item	Oblique Loading
5. No one seems to understand me. (T)	.69
22. Someone has it in for me. (T)	.68
46. I am sure I am being talked about. (T)	.59
20. Much of the time I feel as if I have done something wrong or evil. (T)	.55
45. When in a group of people I have trouble thinking of the right things to talk about. (F)	−.44
27. My conduct is largely controlled by the customs of those about me. (F)	−.43
42. My way of doing things is apt to be misunderstood by others. (T)	.39

This factor appears to be a combination of extreme sensitivity in interpersonal relations, guilt, and ideas of reference. Such a composite suggests the clinical picture of mild paranoia.

Factor III: Social Maladaptation

Item	Oblique Loading
31. I liked school. (F)	.74
10. During one period when I was a youngster I engaged in petty thievery. (T)	.73
23. At school I was sometimes sent to the principal for cutting up. (T)	.72
49. I have never been in trouble with the law. (F)	.42
37. My parents have often objected to the kind of people I went around with. (T)	.41

This is perhaps the only factor which conforms well to a commonly held stereotype of psychopathy, that is, antisocial or asocial behavior.

Factor IV: Impulse Control

Item	Oblique Loading
7. I have had very peculiar and strange experiences. (T)	−.71
50. I have periods in which I feel unusually cheerful without any special reason. (F)	.67

Item		Oblique Loading
38. I like to talk about sex.	(F)	.57
44. Sometimes without any reason or even when things are going wrong I feel excitedly happy, "on top of the world."	(F)	.50
47. I have very few fears compared to my friends.	(F)	.48
42. My way of doing things is apt to be misunderstood by others.	(T)	−.45
4. At times I have very much wanted to leave home.	(T)	−.38
25. At times my thoughts have raced ahead faster than I could speak them.	(F)	.36

The positive pole of this factor (Items 25, 38, 44, 47, 50) suggests a lack of spontaniety in feeling and action. The negative pole (Items 4, 7, 42), conversely, indicates a loosening of impulse controls.

Factor V: Emotional Deprivation

Item		Oblique Loading
36. There is very little love and companionship in my family as compared to other homes.	(T)	.68
29. What others think of me does not bother me.	(F)	.63
26. I believe my home life is as pleasant as that of most people I know.	(F)	.63
16. I do not mind being made fun of.	(F)	.55
34. I wish I were not so shy.	(F)	−.39
13. I wish I could be as happy as others seem to be.	(T)	.38

These items suggest a need for, and also a feeling of having been deprived of, emotional support.

In general, these five factors support some of the rational groupings of items previously noted by McKinley and Hathaway (1956) in their description of the construction of the *Pd* scale: "tendency to respond in overperfect ways" (positive pole of Factor I); "depression and absence of strongly pleasant experiences" (negative pole of Factor I and possibly Factor V); "paranoid trends" (Factor II); and "social trouble" (Factor III). It is significant that the "family trouble" cluster noted by McKinley and Hathaway did not appear as a unitary factor. On the contrary, items relating to conflict with the family (Table 1, Items 4, 11, 18, 26, 36, 37, 39, 43) were well scattered among the several factors. This would suggest that the complexity of family relationships precludes evaluating them on a simple good–bad or trouble–harmony continuum.

Table 2

Correlations Among the Primary Axes

	I	II	III	IV	V
II	−.12				
III	−.17	.10			
IV	.22	−.09	−.22		
V	−.56	.09	−.04	.12	

Table 2 shows the correlations among the primary axes. The largest correlations tend to occur mainly with Factor I (Self-Esteem) and are consistent with the interpretations of the factors. For example, low Self-Esteem people will tend toward Hypersensitivity, Emotional Deprivation, Social Maladaptation, and low Impulse Control.

Factor scores for each *S* were obtained by summing the number of high-loading items[6] on each factor which were answered in the keyed (*Pd*) direction. (Items with negative loadings were scored in reverse.) Table 3 shows the Pearson correlations between these scores and the MMPI *K* scale for a random sample of 50 addicts. Factor items which overlap with *K* have been eliminated in computing these correlations, thereby reducing Factors I, II, IV, and V to seven, six, five, and five items, respectively. Apparently *K* serves as a significant suppressor on each of the factors. Its action on the two bipolars (I and IV), however, is to suppress the negative pole. It may well be that a good deal of the covariance in Table 2 can be attributed to the operation of *K*.

The factor scores used in comparing depressed and nondepressed addicts were computed in the following manner: all *Pd* factor items which overlapped with the *D* scale were initially given a Draw-A-Man IQ by entering the mental age derived from the drawing into the 1937 Stanford-Binet tables (Method A). Next, for *S*s older than 13.0 years, a second IQ was computed by dividing *S*'s MA by Goodenough's recommended maximum CA of 13.0 (Method B). For children under 13 years of age, Goodenough agrees with Terman and Merrill in dividing MA by actual CA, so Methods A and B are identical.

---

[6] Items which are listed under each of the factors (above).

*S*s were divided first into two main groups on the basis of degree of retardation (mild or moderate) as determined by previous individual intelligence tests. Within the major groups, individuals were next classified according to chronological age into eight subgroups. All of the subgroups within each major category were found to be well matched in original IQ.

Results

The analysis of the data preceeded in two general steps, corresponding to the two questions to be answered.

Mental Age

Question 1 inquiries whether Draw-A-Man MA continues to increase after 13.0 years. To answer this question it is necessary to obtain the mean raw score at each of the eight CA levels. As shown in Table 1, the mean raw scores and the MAs which they represent continue to increase fairly rapidly through age 15 in both groups. This increment is contrary to Goodenough's opinion, implied when a maximum CA divisor of 13.0 was established, that performance on the test does not increase beyond that age.

Data in Table 1 is arranged to facilitate inspection of trends. However, in this form it is not amenable to statistical tests of significance. No group averages 13.0 in chronological age, although the logical comparison would seem to be between persons who have just reached Goodenough's ceiling and persons chronologically as far beyond that ceiling as possible. Within each major group, therefore, data were rearranged to provide two age groups consisting of persons age 12–13 (average age 12.9) and persons age 15–16 (average age 15.8). The results show clearly that children increase considerably in ability to produce scorable elements in the human figure after 13.0 years of age. Difference between mean raw scores of the two age groups is significant at the .01 level ($t = 7.58$) for mildly retarded and at the .02 level ($t = 2.38$) for moderately retarded.

Table 1

Mean Performance on the Draw-A-Man Test at Various Chronological Ages

	Moderately Retarded			Mildly Retarded		
Age	*N*	Score	MA	*N*	Score	MA
9[a]	22	4.44	4.11	22	11.22	5.81
10	31	7.00	4.75	35	13.01	6.25
11	24	10.00	5.50	34	14.73	6.68
12	35	10.51	5.63	46	17.34	7.34
13	38	11.59	5.90	42	18.62	7.65
14	24	12.30	6.10	62	19.92	7.98
15	32	13.77	6.44	48	22.86	8.72
16	14	13.41	6.35	27	22.53	8.63

[a] Includes *S*s age 9–0 to 9–11.

One aspect of these results is perplexing. A consistent finding in regard to growth of mental age is that persons age 16 show slightly greater mental maturity than persons age 15. The Binet norms allow four months of MA growth for the CA period from 15.5 to 16.0 in the normal child. Yet, in this study, 16-year-olds do no better than 15-year-olds. A clue may be provided by the reduced number of *S*s at age 16. Contact with several institutions for the mentally retarded indicates that, due to crowded conditions, the average retarded child completes his schooling before age 16. Children retained in school for an extra year would then constitute a selected sample, possibly composed of persons progressing more slowly academically because of social or emotional immaturity (White, 1952). The relationship between lowered Draw-A-Man performance and emotional maladjustment or immaturity is well documented in the literature, and finds widespread application in the Machover (1949) Draw-A-Person test. It seems probable that the present sample at age 16 should not be considered representative of mentally retarded persons in general of that age.

IQ

Although raw score has been shown to increase significantly after Goodenough's maximum CA divisor of 13.0, such a finding is of little importance unless it can also be shown

---

Mitchell for this assistance. Rescoring of the 100 drawings by the writer after one year showed a correlation between scoring and rescoring of .96.

Table 2

Mean IQs on the Draw-A-Man Test at Different Age Levels, Computed by Two Methods

Age	Moderately Retarded A	Moderately Retarded B	Mildly Retarded A	Mildly Retarded B
9	42.22		60.93	
10	45.70		59.66	
11	46.74		57.66	
12	44.98		59.08	
13	45.08	45.74	59.23	60.77
14	43.39	46.78	57.28	61.30
15	43.96	49.89	59.47	67.14
16	42.21	49.16	57.86	66.64

that such increment is sufficient to appreciably raise the IQ. To answer this question, mean IQ's by Methods A and B have been calculated at all CA levels and appear in Table 2. It is assumed here that, ideally, IQs would remain constant over the entire CA range and, furthermore, that CA groups, similar in intelligence, constitute equivalent proxies for a single group of people examined at successive CA levels. Table 2 shows that Method A yields about the same mean IQs at all CA levels in the mildly retarded group. There appears to be a slight decline in IQ after age 14.0 in moderately retarded; however, persons age 16 achieve about the same, and persons age 14 and 15 do better than persons age 9. On the other hand, Method B allows a gradual increase in mean IQ after age 12 in both groups.

To determine the significance of these trends, all children in each major group below CA 13.0 were combined into a single large subgroup, characterized by Draw-A-Man IQs identical whether computed by Method A or B. This large subgroup would constitute a control, the performance of which should be approximated closely by groups of older *S*s similar in individual test IQ. This large subgroup is compared with *S*s aged 15.0 to 16.9. Results suggest that Method A is superior to Method B for the present sample. Method B allows an average increase of 4.25 IQ points for moderately retarded and 7.58 points for mildly retarded patients from age below 13.0 to age 15.8. In fact, Method B shows about half of the present moderately retarded sample age 15 and 16 scoring in the mildly retarded range, a situation clearly out of line with their intellectual capacities as measured by individual intelligence tests. In Method B, differences between the means are significant (for moderately retarded $t = 2.28$, $P < .02$, for mildly retarded $t = 4.92$, $P < .01$). No significant change in mean IQ is found between *S*s younger than 13.0 and *S*s older than 14.9 when Method A is employed ($t = .49$ for moderately retarded and .99 for mildly retarded).

Summary

This study attempts to determine an appropriate maximum CA divisor for the Draw-A-Man test. The sample consisted of 536 mentally retarded children, grouped into eight age levels from 9 to 16 years.

The children in the present sample were found to show growth in Draw-A-Man mental age through at least the sixteenth year. When IQ was computed using the recommended maximum CA divisor of 13.0, the growth in MA after age 13.0 resulted in significant, spurious increases in IQ. When the maximum CA divisor was placed at a higher age level, IQ remained constant over the entire CA range tested.

The results indicate that a maximum CA divisor of 13.0 is inappropriate in calculating Draw-A-Man IQ; the significant MA increment after age 13.0 should be counterbalanced by a corresponding CA increment. An appropriate pattern for CA increase through the years from 13 to 16 was found to be that used by Terman and Merrill in constructing the Stanford-Binet IQ tables.

REFERENCES

BIRCH, J. W. The Goodenough Drawing Test and older mentally retarded children. *Amer. J. ment. Defic.*, 1949, **54**, 218–224.

GOODENOUGH, FLORENCE L. *Measurement of intelligence by drawings*. New York: World Book, 1926.

MACHOVER, KAREN. *Personality projection in the drawing of the human figure*. Springfield, Ill.: Charles C Thomas, 1949.

TERMAN, L. M., & MERRILL, MAUD A. *Measuring intelligence*. Cambridge: Houghton Mifflin, 1937.

WALLIN, J. E. W. *Children with mental and physical handicaps*. New York: Prentice-Hall, 1949.

WECHSLER, D. *Measurement of adult intelligence*. Baltimore: Williams & Wilkins, 1944.

WHITE, W. D. The education and training program for institutional care of the mentally handicapped. *Amer. J. ment. Defic.*, 1952, **56**, 765–770.

(Received January 5, 1959)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

A NOTE ON "SOCIAL DESIRABILITY AS A VARIABLE IN THE EDWARDS PERSONAL PREFERENCE SCHEDULE"

ALLEN L. EDWARDS, CALVIN E. WRIGHT, AND CLIFFORD E. LUNNEBORG
University of Washington

The Edwards Personal Preference Schedule (EPPS) is a forced-choice inventory in which, with standard instructions, subjects are asked to describe themselves by choosing between AB pairs of personality statements.

A group of 134 psychology students at the University of Washington was given the *complete* EPPS with standard instructions to describe themselves. Another group of 114 students was given the *complete* EPPS with instructions to choose the alternative they considered to be more socially desirable. A third group consisted of a sample of 160 records drawn from the normative group used in standardizing the EPPS. These three groups will be referred to as the W, SD, and N groups, respectively.

For each group, the proportion giving the A response to each of the pairs of statements was found separately for males and females. These proportions were then weighted to provide estimates of those to be expected from a group in which the ratio of males to females is 87/54. The reason for weighting the item statistics was to make them comparable to the weighted item statistics of another study done at the University of Buffalo (Corah, Feldman, Cohen, Gruen, Meadow, & Ringwall, 1958).

The Buffalo study and the present one are similar, except for the fact that in the Buffalo investigation the item statistics were obtained only for a *selected set of 30 items* which were *administered out of context* of the EPPS. Since there is other evidence to support the belief that lifting items out of context of a standardized test may influence the nature of the items and responses to them, we wished to investigate this possibility with respect to the results reported in the Buffalo study.

To determine the similarity between the item statistics for the W and N groups, the proportions choosing A in each item were correlated. For the complete set of 210 items, the correlation was .93. For the 30 items used in the Buffalo study, the correlation was .91.

We then found the correlation between the item statistics of the SD group and those of the W and N groups. For the complete set of 210 items these correlations were .69 and .61, respectively. For the selected set of 30 items used in the Buffalo study, the two correlations were .64 and .51, respectively.

The results reported above are not consistent with those obtained in the Buffalo study. In the Buffalo study, a correlation of .88 was found between the proportions choosing A under standard instructions and under instructions to choose the more socially desirable alternative for the selected set of 30 items. On the other hand, the correlations we obtained for the same set of 30 items were .64 (between SD and W) and .51 (between SD and N). The two investigations differ in that in the Buffalo study, the 30 items were, for both sets of instructions, administered out of context, whereas in our study they were not.

The results we obtained serve to emphasize, once again, the fact that removing items from the context of a standardized test may alter the nature of the items and responses to them.

REFERENCE

Corah, N. L., Feldman, M. J., Cohen, I. S., Gruen, W., Meadow, A., & Ringwall, E. A. Social desirability as a variable in the Edwards Personal Preference Schedule. *J. consult. Psychol.*, 1958, *22*, 70–72.

(Received October 8, 1958)

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

BRIEF REPORTS

BENDER-GESTALT ROTATIONS AND EEG ABNORMALITIES IN CHILDREN[1]

SHERWOOD B. CHOROST, GEORGE SPIVACK, AND MURRAY LEVINE
Devereux Foundation, Devon, Pennsylvania

Hanvik (1954) has reported that 80% of children who rotated one or more Bender figures had abnormal EEGs. He concludes that "the rotation of the figures of the Bender-Gestalt test is even more highly predictive of brain damage in children than among adult patients" (p. 399).

Our purpose was to follow up Hanvik's suggestive findings with children, since he presents no control data nor makes reference to the antecedent probability of abnormal EEGs in his population.

Our population consisted of 68 children, both male and female, under the age of 18, in a residential treatment center.

The *S*s were selected, having met two criteria: (*a*) an EEG examination and (*b*) a scorable Bender-Gestalt protocol. The Bender is administered routinely, but EEGs were available *only* if there had been medical indication for this procedure. Hanvik does not indicate whether Benders and EEGs are administered routinely in his clinic.

The Bender-Gestalt records were scored for rotations of 30° or more from the normal axis, following Hanvik and Anderson (1950). Of those children who had one or more Bender rotations ($N = 51$), 69% had abnormal EEGs. This approaches Hanvik's findings of 80%. Of those with no rotation ($N = 17$), only 47% had abnormal EEGs. This difference is significant ($p = .05$).

The probability of obtaining an abnormal EEG in our sample is 63%. Using the presence or absence of one or more Bender rotations as a predictor of abnormal EEGs, we could make the correct decision in 65% of the cases. Using Bender rotations does not increase our predictive efficiency.

Hanvik's (1954) conclusion that Bender rotations are highly diagnostic of brain damage in children is supported neither by his study nor by ours. He reports neither base rates of EEG abnormality nor the frequency of predictive failures which would enable us to evaluate the diagnostic efficiency. Hanvik did not report on the number of cases that had no rotation and that also had EEG abnormality. Unless EEGs are given routinely, the frequency of EEG abnormalities is likely to be very high, and the correlation he reports could reflect this.

Our data point up the importance of assessing the diagnostic utility of any clinical tool against some known base rate. However, there is an important distinction to be drawn between diagnostic efficiency and scientific validity. If the EEG criterion of brain damage has some validity despite all the recognized weaknesses, individuals who do rotate Bender designs are more likely to have brain damage than those who do not rotate. We cannot use this fact diagnostically with any efficiency. However, lest our newly found understanding of the base rate problem in diagnosis discourage us, it should be pointed out that the obtained results are still meaningful in helping us to note some of the perceptual and psychomotor consequences of brain damage.

REFERENCES

HANVIK, L. J. A note on rotations in the Bender-Gestalt Test as predictors of EEG abnormalities in children. *J. clin. Psychol.*, 1954, **9**, 399.

HANVIK, L. J., & ANDERSON, A. L. The effect of focal brain lesions on recall and on the production of rotations in the Bender-Gestalt Test. *J. consult. Psychol.*, 1950, **14**, 197–198.

(Received April 24, 1959)

---

[1] An extended report of this study may be obtained without charge from Sherwood B. Chorost, Devereux Schools, Devon, Pa., or for a fee from the American Documentation Institute. Order Document No. 6021, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

EYE–EAR EMPHASIS IN THE DAP AS INDICATING IDEAS OF REFERENCE[1]

ALBERT V. GRIFFITH AND D. A. R. PEYMAN[2]
Bryce Hospital, Tuscaloosa, Alabama

The purpose of this study was to relate eye-ear emphasis in the DAP with the trait, "ideas of reference." The *S*s were drawn from 745 male patients who had consecutively entered Bryce Hospital. The DAP drawings of this population were jointly scanned by two staff psychologists. Eighteen *S*s were selected as emphasizing eyes and/or ears. Only those *S*s were selected when both judges were in agreement as to the presence of the sign. A conservative position was taken; i.e., only those *S*s were chosen where eye and/or ear emphasis was obvious. An 8% sample ($N = 58$), which was to serve as the control group, was then randomly selected from the remaining population.

The 76 hospital records (eye-ear emphasis $N = 18$; control $N = 58$) were randomly presented to two other staff psychologists. Judgments were jointly made by the latter as to the presence of ideas of reference for each *S*. These judgments were not based on any psychometric data. As the two judges worked together, no estimate of reliability was made. Both judges felt that it was much easier to judge the presence of ideas of reference than to decide in what diagnostic category each *S* should be placed.

Eleven of the 18 *S*s in the eye-ear emphasis group were judged as having ideas of reference, while 13 of the 58 *S*s in the control group were so judged. A statistical analysis of the data indicated that a significant difference existed between the two groups ($\chi^2 = 7.754$ for one *df*; $P < .01$). It is concluded that the hypothesis is confirmed.

Swensen's judgment (1957, p. 461) that the DAP is of use in clinical work only as a rough screening device was based on studies that had certain defects of design. Most of these studies used psychiatric diagnoses as the criteria. The reliability of psychiatric diagnoses, however, has been questioned for years. If psychiatric diagnosis (paranoid involvement) had been the criterion in the present study, nonsignificant results would have been obtained. Only 5 of the 18 *S*s in the eye-ear emphasis group were diagnosed as having paranoid involvement, while 8 *S*s in the control group were so diagnosed.

What was being tested in most of the studies reviewed by Swensen was the efficiency of DAP signs in predicting various psychiatric labels. The most important thing for the clinician, however, is the validity of the sign; i.e., *when the sign is present*, does it signify beyond chance expectancy a given condition or trait. The present study was designed to investigate this type of problem. The result was that eye-ear emphasis in the DAP was demonstrated to be predictive of ideas of reference.

Our data indicate that eye-ear emphasis in the DAP, while valid, is inefficient in predicting ideas of reference. This should not surprise us, as it is generally recognized that tests are inefficient in predicting personality variables.

REFERENCE

SWENSEN, C. H., JR. Empirical evaluations of human figure drawings. *Psychol. Bull.*, 1957, **54**, 431–466.

(Received April 27, 1959)

---

[1] An extended report of this study may be obtained without charge from D. A. R. Peyman, Psychology Department, Bryce Hospital, Station 3, Tuscaloosa, Alabama, or for a fee from the American Documentation Institute. Order Document No. 6022, remitting $1.25 for microfilm or $1.25 for photocopies.

[2] The authors wish to thank Clem C. Nesmith and Waters C. Paul, staff psychologists at Bryce Hospital, for serving as judges in the present study.

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

THE VALIDITY OF THE COLUMBIA MENTAL MATURITY SCALE[1]

BETSY WORTH ESTES, FRANK KODMAN, AND MACY AKEL
University of Kentucky

Previous studies of the 1954 Columbia Mental Maturity Scale (CMMS) indicate that further investigation is warranted before uncritical use is made of it with motor damaged and/or verbally handicapped children. The present study uses the Wechsler Intelligence Scale for Children (WISC) as the criterion.

The test sample consisted of 25 male and 25 female elementary school children in a middle-class suburban school in Kentucky. The age range was 7–2 to 9–4 years with a mean age of 8–2 years. The study was designed so that the sample would have a mean and standard deviation which would not differ from the CMMS standardization group for ages seven, eight, and nine years (Akel, 1958). The obtained t between the present sample and the standardizing sample equalled 1.05, with 98 *df*. An F test indicated that the variances were homogeneous. Further details may be found in Akel (1958).

The CMMS was correlated with the WISC Full Scale (FS), Verbal Scale (VS), Performance Scale (PS), and the WISC subtests. A correlation matrix was constructed, and Osgood and Suci's method of factor analysis was used (1955).

Coefficients of correlation between the CMMS and WISC FS, VS, and PS, respectively, were .49, .53, .36. Correlations between CMMS and WISC subtests ranged from .05 to .51. Thus the correlations range from very low to moderate only. The use of Fisher's z transformation demonstrated that the only correlation which differed significantly from that of the CMMS and WISC FS was Picture Completion.

Factor analysis of the correlation matrix indicated five factors. All tests were loaded on Factor I, which may represent general intelligence. The other four factors were specific to certain subtests: Factor II, Object Assembly; III, the CMMS and Picture Completion; IV, Coding and Picture Arrangement; and V, Similarities. A larger sample is needed to assess the validity of the factor loadings.

The validity of the CMMS is not upheld by the present study. The CMMS should not be used as a substitute for intelligence tests other than the Stanford-Binet, although it may have limited use when other tests cannot be used.

REFERENCES

AKEL, MACY. An investigation of the relationship between the Columbia Mental Maturity Scale and the WISC. Unpublished master's thesis, Univer. of Kentucky, 1958.

OSGOOD, C. E., & SUCI, C. J. Factor analysis of meaning. *J. exp. Psychol.,* 1955, **50**, 325–338.

(Received April 22, 1959)

---

[1] An extended report of this study may be obtained without charge from Frank Kodman, Jr. or Betsy Worth Estes, Department of Psychology, University of Kentucky, Lexington, Kentucky, or for a fee from the American Documentation Institute. Order Document No. 6020, remitting $1.25 for microfilm or $1.25 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

PERSONALITY FACTORS IN NURSERY SCHOOL CHILDREN AS DERIVED FROM TEACHER RATINGS[1, 2]

DONALD R. PETERSON AND RAYMOND B. CATTELL
University of Illinois

This study is part of a larger research program designed to provide descriptive information about personality structure in children and young adults (Cattell, 1957). Cross-sectional assessment of personality at several different age levels has been effected by observing and rating behavior *in situ*, eliciting responses to questionnaires and administering a variety of objective tests. Data within each medium at each age level have been separately factorized, and the results compared across media and over age.

The present study deals with teacher ratings of preschool children. Eighty pupils in a metropolitan university nursery school were ranked by their teachers on 36 variables representing spaced aspects of the personality sphere. Pooled judgments were intercorrelated, a centroid factor analysis was conducted, and rotations to oblique simple structure were performed. Most of the factors previously isolated in older subjects were found here. Nine factors (in Cattell's terminology, C, H, E, G, K, F, I, L, and A) were identified with reasonable confidence. Four others were tentatively identified as D, J, M, and O. A recently stated proposition (Cattell, 1957) that the structure of personality undergoes no radical change during childhood was thus confirmed and the finding extended to early childhood.

Although the same factors evidently emerged throughout the age range examined in the series of investigations of which this study is a part, factor composition sometimes appeared to change in a systematic way. Two such apparent changes, which might form hypothetical bases for future research, may be summarized as follows. Proneness to Neuroticism vs. Ego Strength (C−) exhibits a developmental change from uncomplicated emotional instability and sensitivity in early childhood, through predominant expression as conduct disorder in middle childhood, to predominant expression as neurotic symptomatology among adults. Super Ego Strength (G) develops out of pure sensitivity to emotion-inducing stimuli in early childhood. This diffuse sensitivity is followed by the emergence of positive, initiatory super ego tendencies, and only after these have developed do the negative inhibitory aspects of super ego evolve.

REFERENCE

CATTELL, R. B. *Personality and motivation structure and measurement*. Yonkers, N. Y.: World Book, 1957.

(Received April 17, 1959)

---

[1] This project was supported by the Graduate Research Board, University of Illinois. The writers wish to express their appreciation to Peter V. Vygantas and Donald Tollefson for their help in gathering the data and analyzing results. We are deeply grateful to the staff of the University of Chicago Nursery Schools, and the Cooperative Nursery Schools of Urbana, for their help throughout the study.

[2] An extended report of this study may be obtained without charge from D. R. Peterson, Psychology Department, University of Illinois, Urbana, Illinois, or for a fee from the American Documentation Institute. Order Document No. 6019, remitting $1.75 for microfilm or $2.50 for photocopies.

Journal of Consulting Psychology
Vol. 23, No. 6, 1959

PSYCHOLOGICAL TEST REVIEWS

New Tests

Weingarten, Kurt P. *Picture Interest Inventory.* Grades 7 to adult. Untimed, (35) min. Test booklet; answer sheets; scoring keys; manual; pp. 24. Los Angeles: California Test Bureau, 1958.

The *Picture Interest Inventory* (PII) makes use of simplified line drawings of men engaged in occupational activities. It consists of two parts: Part I contains 53 groups of three pictures each to which the subject responds by choosing the represented activity he likes the most and the least; Part II consists of 30 additional pictures each of which is responded to separately. Six occupational scores (Interpersonal Service, Natural, Mechanical, Business, Esthetic, Scientific) and three nonoccupational (Verbal, Computational, Time perspective) are obtained.

Though it rests on a number of studies, only a start has been made on the development of this test. Norms are based on 1,000 males, drawn in unspecified portions from junior high school, high school, and college, and on 140 females (66 12th graders and 74 eighth graders), the latter set being anchored to the norms for the Occupational Interest Inventory. The manual reports that norms for different age groups (males) were within two raw score points in all parts of the distribution for seven of the nine scales and the other two differed by three points in only a few points of their respective distributions. No basis is given for judging comparability of age samples with respect to such factors as intelligence of subjects or the social class of their parents. Reliabilities are based on test–retest with one week interval and are acceptable, though not outstanding. The major claim for the test's validity rests on substantial correlations with appropriate scale of the *Occupational Interest Inventory,* moderate correlations with the *Kuder Preference Record,* and highly varied correlations with the *Vocational Interest Blank.* The only reported study of females reveals the presence of distinct sex differences. Though much emphasis is placed on the test's nonverbal character and its consequently greater applicability to 7th graders, neither longitudinal data nor data comparing its effectiveness with verbal tests are provided.

This may prove to be a very useful addition to the kit of vocational interest tests. Much more work needs to be done before it can be accepted with confidence.—E. S. B.

Eysenck, H. J. *Maudsley Personality Inventory.* Adolescent and adult. Test booklet (4 d); scoring keys (11 each); manual, pp. 8 (2/6). London, England: University of London Press LTD, 1959.

This brief questionnaire of 48 items and its even briefer short form (12 items) have played an integral part in the author's well known research on personality. Many instruments have been launched for full scale use with much less behind them. The manual represents the height of English diffidence. Only the briefest summary is given of a few of the salient results of research and the reader is referred to the relevant publications. He is told that "the M.P.I. should be regarded as a *research instrument.* Different firms, organizations, hospitals, universities, and other bodies have different problems, deal with different samples of the population, and aim at different solutions of their problems. Only applied research can determine whether instruments such as the M.P.I. can be successfully used by them, and just what form such use can best take." No high powered American merchandising here!—*E. S. B.*

Technical Report on the California Test of Mental Maturity; pp. 32. Los Angeles: California Test Bureau, 1957.

This report is intended as a comprehensive summary of the development of the California Test of Mental Maturity in its various editions and short forms. It is designed to supplement the manuals for the six levels of the two series (short and long form). Extensive treatment is given to history and standardization. Validity data are confined to correlations with other pencil-and-paper intelligence tests. Contrary to discussions of construct validity in *Technical Recommendations For Test Manuals,* it is claimed that no other kinds of validating data are applicable.—*E. S. B.*

Andrew, Dorothy M., & Patterson, D. G. *Manual for Minnesota Clerical Test;* pp. 12. New York: Psychological Corporation, 1959.

The previous revision of this well known test was issued in 1946. Approximately six new studies are reported; the results are much in the same pattern as earlier research. A much wider selection of norm groups is available in this manual.—*E. S. B.*

Jenkins, T. N. *How Well Do You Know Yourself?* (Form NF-21) Adolescent and adult. Specimen Kit ($2.50 each) containing test booklet (with attached profile sheet), scoring keys, and a manual, pp. 26. New York: Executive Analysis Corporation, 1959.

This 120-item inventory yields 17 trait scores, e.g., Irritability, Novelty Loving, Vocational Assurance, Submissiveness, Impulsiveness, and two scores regarding test taking attitudes, Consistency and Objectivity. The trait scores are founded on factor analyses. Consistency is simply the tendency to give the same answer to the same item at two different times. The test objectivity items are a refinement of a set empirically derived by differentiating two groups of subjects who, on the basis of interviews, were judged to differ in the realism of their conceptions of themselves. The general format of the test and manual leave something to be desired. The test booklet incorporates a profile and answer sheet and, thus, is expendable, which naturally increases the cost of using this instrument. The manual is written in a pedantic, sometimes opaque style.

This test may prove equal or superior to other personality inventories, but the case is not proved by the data summarized in the manual. Although satisfactorily high multiple correlations are reported for the six items of each factor with that factor, the test–retest correlations, not unexpectedly, are low in too many instances. Norms for secondary school, college undergraduate, and office and factory populations of both sexes are based on small inadequately defined samples. While there is evidence of considerable internal analysis of the test, external validation is limited to three studies, all doctoral theses, only one of which has been reported in the general literature. Much more needs to be done.—*E. S. B.*

Strong, Jr., E. K. *Manual for Strong Vocational Interest Blanks;* pp. 40. Palo Alto, Calif.: Consulting Psychologists Press, 1959.

This is an expanded manual which takes the place of separate manuals for the Men's and Women's forms of this most highly respected interest inventory. More attention than in previous versions is given to the interpretation of test scores. A "Publisher's Foreword" rightfully calls attention to the volumes of relevant research and writings which cannot be fully presented in a manual and recommends three additional books to "the serious student of interest measurement and the professional counselor."—*E. S. B.*

ERRATUM

The review of the *Employee Aptitude Survey* in the June issue of this Journal (p. 284) contained the misstatement that "In a few instances no data on the correlations between alternate forms is offered." One of these instances turned out to be Test 9 which has no alternate form. This error served as a false support for a reading error in failing to note that the manual for Test 8 listed the alternate form correlation in a context different from the manuals for other tests in the battery.